Katholieke Universiteit Leuven
Faculteit Psychologie en Pedagogische Wetenschappen

Centrum voor Leerpsychologie en Experimentele Psychopathologie

Changing and assessing valence

Proefschrift aangeboden tot het verkrijgen van de graad van
Doctor in de Psychologie
door
Inneke Kerkhof

o.l.v.
Prof. Dr. Dirk Hermans, promotor
Prof. Dr. Frank Baeyens, copromotor
Prof. Dr. Debora Vansteenwegen, copromotor

2010

Inneke Kerkhof
Changing and assessing valence
Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologie, 2010
Prof. Dr. Dirk Hermans (promotor), Prof. Dr. Debora Vansteenwegen (copromotor) en Prof. Dr. Frank Baeyens (copromotor)

Het fenomeen van terugkeer van vrees na een succesvolle exposure behandeling biedt een uitdaging voor onderzoek en praktijk. Het huidige doctoraatsonderzoek bekijkt deze problematiek vanuit een leerpsychologische visie op angst(stoornissen). Volgens conditioneringstheorieën leidt de contingente aanbieding van een voorheen neutrale prikkel (de geconditioneerde prikkel of CS genoemd) met een bedreigende gebeurtenis (de ongeconditioneerde prikkel of US) tot de vorming van een associatie tussen de geheugenrepresentaties van beide stimuli. Latere confrontatie met de CS zal de US-representatie opnieuw activeren, alsook de vrees die er mee geassocieerd was. Vanuit dit perspectief kunnen exposure en terugkeer van vrees beschouwd worden als het klinisch analoog van respectievelijk extinctie (d.i. onbekrachtigde aanbiedingen van de CS na acquisitie) en de terugkeer van vreesresponsen na extinctie. Een belangrijk inzicht uit vreesonderzoek is dat doorheen een fase van vreesconditionering de CS niet alleen een voorspeller wordt voor de bedreigende US - en derhalve vrees ontlokt, maar tevens voorzien wordt van een negatieve valentie. Deze evaluatief geconditioneerde negatieve valentie blijkt relatief weerstandig te zijn aan uitdoving en zou een affectief-motivationele bron kunnen zijn voor de terugkeer van vrees. Verschillende studies vonden inderdaad dat deze overgebleven negatieve valentie na extinctie voorspellend was voor de mate van terugkeer van vrees. Deze bevindingen suggereren dat mensen met een angststoornis beter gebaat zouden kunnen zijn met een behandeling die niet enkel focust op exposure, maar die ook inwerkt op de geconditioneerde negatieve valentie van het vreesobject. Verrassend genoeg is er nog maar weinig onderzoek verricht naar mogelijke technieken waarmee geconditioneerde valentie gewijzigd kan worden. Het huidige doctoraatsonderzoek probeerde hieraan tegemoet te komen.

In een <u>eerste onderzoekslijn</u> werd onderzocht of geconditioneerde valentie gewijzigd kan worden met een contraconditionerings (cc) procedure (d.i. het paren van de CS met een nieuwe US met een evaluatieve waarde die tegenovergesteld is aan deze van de oorspronkelijke acquisitie US). Vijf studies werden uitgevoerd binnen deze onderzoekslijn. In een eerste stap (experiment 1, 2) gingen we op zoek naar een goed evaluatief conditioneringsparadigma waarin we deelnemers nieuwe voor- en afkeuren zouden kunnen aanleren. Succesvolle resultaten werden behaald met een paradigma waarin foto's gepaard werden met het eten van lekkere of slecht smakende koekjes. In een tweede stap (experiment 3) onderzochten we in dit paradigma of nieuw aangeleerde preferenties gewijzigd kunnen worden via cc. Dit bleek het geval te zijn. In een derde stap (experiment 3, 4) onderzochten we de duurzaamheid van valentiewijzigingen ten gevolge van cc door na te gaan of de veranderde valentie intact bleef na verloop van tijd of na een contextverandering. Ook dit bleek zo te zijn. Onze bevindingen lijken er dus op te wijzen dat cc een beloftevolle techniek is om eerder geconditioneerde valentie op een duurzame manier te wijzigen. In een laatste stap (experiment 5) ten slotte, onderzochten we in een muizenstudie of terugkeer van vrees gereduceerd kan worden door cc toe te passen na extinctie. We vonden hier geen evidentie voor, maar aangezien geen valentiemeting werd opgenomen in deze studie kunnen we niet uitsluiten dat onze cc manipulatie mogelijk niet sterk genoeg was.

In een <u>tweede onderzoekslijn</u> focusten we op het meten van valentie. Indirecte reactietijdtaken zijn erg populair in valentieonderzoek omdat ze gemakkelijk in gebruik zijn en minder gevoelig zijn voor sociale wenselijkheidseffecten dan ratingschalen. Recent worden deze reactietijdtaken steeds vaker gebruikt als voor- en/of nameting in onderzoek dat nagaat hoe evaluaties verworven en veranderd kunnen worden. In dit tweede onderzoeksluik bespreken we de nadelen van zulke voor-/nameting studies en argumenteren we dat het soms beter kan zijn om evaluaties te meten *tijdens* de leerfase ('online'). De bestaande indirecte reactietijdtaken lenen zich echter niet tot een online afname. In het tweede deel van dit doctoraat presenteren we drie studies waarin we met succes een indirecte reactietijdtaak ontwikkelden die *wel* toelaat valentie te meten tijdens de leerfase.

Inneke Kerkhof

Changing and assessing valence

Proefschrift aangeboden tot het verkrijgen van de graad van Doctor in de Psychologie, 2010
Prof. Dr. Dirk Hermans (promotor), Prof. Dr. Debora Vansteenwegen (copromotor) en Prof. Dr. Frank Baeyens (copromotor)

Return of fear after successful exposure treatment is a common finding and constitutes a challenge for clinical practice and fear research. In the current doctoral project, we look at fear from a learning theoretical perspective. According to contemporary learning theories, fear can be acquired through an associative learning process. In a fear conditioning procedure, a neutral stimulus (the conditioned stimulus or CS) comes to evoke fear after it is repeatedly paired with an aversive stimulus (the unconditioned stimulus or US). From this perspective, extinction - a decrease in conditioned responding to the CS due to repeated CS alone presentations - can be viewed as an experimental model for exposure therapy. Return of conditioned responses after extinction can then be seen as a model for relapse after treatment. An important insight for the current project is that during a fear acquisition procedure the meaning of the CS is altered in two important ways. The CS not only becomes a valid predictor for the US, but also acquires a negative connotation through evaluative conditioning that is (relatively) unaffected by an extinction procedure. Several findings indicate that this remaining negative valence after extinction might not be without consequences and might function as an affective-motivational source for return of fear. Different studies, for instance, found the remaining negative valence after extinction to be predictive for the amount of return of fear. These findings suggest that patients with a fear disorder might benefit from a treatment that not only focuses on the disconfirmation of expectancies through exposure, but also targets the acquired negative valence of the fear object. Surprisingly, hitherto little research has been conducted on how conditioned preferences, once acquired, can be altered. The present dissertation is aimed at addressing this need.

In a <u>first line</u> of research we examined whether conditioned valence can be changed through counterconditioning (cc) (i.e., pairing the CS with a new US, which has a valence opposite to that of the original acquisition US). Five studies were conducted in this line. The first two studies (experiment 1, 2) were aimed at finding a robust evaluative conditioning paradigm in which we could study the acquisition and change of conditioned (dis)likes. Based on the results of these first two experiments, we chose to work with a picture-flavor paradigm in the following studies. In experiment 3, we applied this paradigm to examine whether newly acquired preferences could be changed with a cc procedure. This turned out to be the case. In a next step (experiment 3, 4) we investigated the sustainability of cc effects. More specifically, we tested whether counterconditioned preferences remained intact over time and after a context switch. This was found to be so. Our findings thus seem to suggest that cc provides a promising approach for changing previously acquired conditioned valence in a durable way. Finally, in a fifth experiment, we examined in a mice study whether return of fear could be reduced by presenting cc trials after extinction. No beneficial effect was found of the cc treatment. However, as CS valence was not measured in this experiment, we cannot exclude the possibility that our cc manipulation was perhaps not strong enough to change the CS's acquired negative valence.

A <u>second line</u> of research focused on the assessment of valence. Indirect reaction time (RT) measures are popular in evaluation research due to their easy application and the fact that they are assumed to be less vulnerable to demand effects than rating scales. In studies that examine how evaluations can be acquired and changed, indirect RT measures are typically administered in a session that precedes and/or follows the evaluative learning phase (i.e., as a pre-test and/or post-test). In this second line, we discuss the disadvantages of such pre-test/post-test designs and argue that it sometimes might be preferable to assess valence *during* the learning phase (i.e., online). The currently existing RT measures, however, do not lend themselves well to be integrated in an ongoing evaluative learning procedure. In the second part of this dissertation, we present three experiments in which we successfully developed an indirect RT task that *can* be used to assess valence online.

Bedankt

Eindelijk... hier heb ik zo lang naar uitgekeken! Na vier jaar en 2 maanden ben ik er dan toch geraakt, mijn boekje is af!!! Vier rollercoaster jaren die ik niet was doorgekomen zonder de steun en hulp van een heleboel mensen die steeds in mij bleven geloven. Hoog tijd voor een woord van dank dus!

Dirk, je bent echt een promotor uit de duizend. Mijn doctoraat was er eentje met veel ups, maar ook veel downs en die hebben we samen doorgesparteld. Ik apprecieer het heel erg dat ik niet alleen met goed nieuws bij je terecht kon (eerste artikel binnen! *Only just*, maar toch!), maar even welkom was op de momenten dat ik in een dip(je) zat (lang leve tissue dozen!). Bedankt voor alle tijd die je voor mij maakte (de vele woensdagochtenden!), je continue aanmoedigingen, je geduld als ik je bleef stalken met vragen (maar hoe past Adriaan zijn theorie nu juist op mijn data?) of weer eens een artikel of coverletter niet uit handen durfde geven (zou je nog een *allerlaatste* keer willen nakijken of het zo duidelijk is wat ik geschreven heb?), voor het telkens opnieuw vinden van iets positiefs in data die mij wanhopig maakten (weer eens significant in de foute richting, zucht), voor je vertrouwen en de vrijheid die ik kreeg om met mijn project een andere weg in te gaan dan oorspronkelijk gepland was. Het gouden promotor transparantje zou je zeker ook winnen! Bedankt voor alles!!!

Frank, bij jou is mijn CLEP verhaal begonnen. Na de eerste les leerpsychologie van Paul Eelen was ik overtuigd: hier moest ik meer over weten! Ik besloot van thesisonderwerp te veranderen en kwam met knikkende knieën vragen of ik bij jou rond occasion setting mocht komen werken. 'Weer zo'n thesis-hopper' dacht je waarschijnlijk en je stuurde me met drie lange artikels vol FP's, X→A+'en en A-'en naar huis om te zien of ik het nog allemaal wel zo interessant vond na deze lectuur. Het begin van een leerzame en leuke samenwerking die zich voortzette tijdens mijn doctoraatsperiode. Bedankt voor je support along de way: de vele tips en suggesties bij mijn designs, de feedback op mijn schrijfsels, de peptalk in je mailtjes,... maar vooral ook voor je eeuwige vrolijkheid en spontaniteit waarmee je elke CLEP vergadering een plezier maakt.

Deb, op de (regelmatige) momenten dat ik de bomen door het bos niet meer zag en geen structuur kreeg in mijn ideeën of schrijfsels, was ik altijd blij dat ik bij jou terecht kon. Om daarna weer met een hoofd vol ideeën en goede moed verder te werken. Ik bewonder je passie voor onderzoek, je scherpe kritische blik, je recht-door-zee mentaliteit,... Een straffe CLEP madam en een fantastische copromotor. Heel erg bedankt!

Dan de overige CLEPpers: Adriaan, An, Ann(eke), Bart, Bram, Carlos, Corry, Dinska, Els(ie), Elise, Eva, Evelien (Coppie), Fernando, Filip R., Filip V., Hanne, Hans, Heleen, Jeroen, Jorien, Kathleen, Kris, Marie Louise, Marleen, Mathijs, Mathilde, Mieke, Netje, Priya, Riet, Sabine, Tom, Valerie, Yannick(ske): the best colleagues in the world!!!
Bedankt allemaal voor de geweldige sfeer in het PSI, de gezellige middaglunches aan de tafel beneden, de leuke retraites, de kickerwedstrijdjes (ook al durven sommige mensen in het heetst van de strijd wel eens overgaan tot vreemde koosnaampjes), de vele feestjes in de foute Oase, de dagelijkse tetters bij het theehoekje in de 02.113, het samen plezier maken op leuke congressen (Is het een vogel, is het een vliegtuig, nee dat is het nieieieieiet!!! En blijkt nu toch niet net vandaag Tom jarig te zijn!), het samen aftellen naar het einde van een congres in de meest saaie stad van Kroatië, de lieve woorden en support-mailtjes tijdens mijn vermoeiende laatste doctoraatsweken, de kook- en schminkcursussen (dringend tijd trouwens voor het organiseren van breicursus II nu ik terug een leven met tijd heb ☺), de singstar duetten, de troostende woorden na een journal reject of weer eens slechte data, de discussies over het laatste Gawronski model, leningen of ventilatiesystemen, het samen aerobiccen (En... pony, pony!) en zwemmen, de housewarmings, de vele peptalk, de verkleedpartijtjes bij doctoraatsverdedigingen, de verjaardagscadeautjes onder de ladies, de giraf knuffel, de omkoop-snoepjes na een spin exposure, het samen kwetteren op de bureau,... en vooral om zoveel meer te zijn dan gewoon collega's!!! Ik ga jullie zo hard missen, kon ik jullie maar meenemen naar mijn volgende werkplek(ken)...
Special thanks aan mijn rookie-genootjes: jullie zijn één voor één gewoon geweldig! Ik heb genoten van onze meetings en vele zotte momenten samen. Heel erg bedankt ook voor jullie naleeswerk en lieve schrijfsels bij mijn laatste doctoraatsteksten (natuurlijk heb ik ook toen mijn zakdoek moeten bovenhalen ☺). Die koekjes komen er nu echt wel gauw aan! En ten slotte nog een extra mercike aan Dinska omdat ik haar doctoraatsboekje als lay-out voorbeeld mocht gebruiken.

Ook de LBP groep wil ik heel erg bedanken voor de fijne momenten in de kelders van het PSI. In mijn eerste doctoraatsmaanden heb ik meer tijd bij jullie doorgebracht dan in het CLEP. Bedankt voor de tips bij mijn (eeuwigdurende) muizenexperiment, de toffe babbels tijdens het afwassen van de kooien, het samen wegen en staartjes kleuren, het samen shocks krijgen bij het op zoek gaan naar die ene staaf in de grid die niet meer goed werkte, de bowling uitjes,... alsook de leuke retraites (met bijhorende quiz en feestjes) samen! Jullie zijn echt een toffe bende!

Verder wil ik ook Elfi Goesaert en Nina Van den Bussche bedanken die in het kader van hun stage en thesis meehielpen aan mijn onderzoek. Bedankt voor jullie engagement en enthousiasme, ik vond het heel leuk om met jullie samen te werken!

Ik wil ook de leden van mijn begeleidingscommissie bedanken voor hun interesse in mijn onderzoek en hun deskundige commentaren: Prof. Dr. Robert Boakes, Prof. Dr. Jan De Houwer, Prof. Dr. Rudi D’Hooge, Prof. Dr. Merel Kindt, Prof. Dr. Batja Mesquita en Prof. Dr. Ilse Van Diest.

Een grote merci ook aan mijn lieve vrienden, in het bijzonder aan Vanes, Leen, Erik (& co) en de theo-bende: bedankt voor de aanmoedigingen, de afleiding, dat jullie naar Leuven kwamen wanneer ik tot over mijn oren in de schrijfsels zat, de etentjes, het tegen elkaar opklagen over doctoraatsdips, het samen lachen met de PhD comics, de geweldige trouwfeesten (hou jullie klaar voor 31 juli!), jullie geduld tijdens deze laatste maanden waarin ik zo weinig van mij liet horen... Joa, extra merci dat ik altijd bij jou terecht kon met C++ programma’s die niet deden wat ik wou, dubieuze Engelse zinnen en artikels die weer eens 300 woorden te lang waren. Bedankt voor alle peptalk en het doorsturen van de vele comics die mijn dag steeds weer goed maakten. Kom binnenkort nog maar eens Brel zingen bij ons ☺

Ook mijn lieve familie en schoonfamilie wil ik bedanken voor vier jaar lang trouw supporteren. Ik heb het zo getroffen met jullie... No time om over doctoraatsissues te stressen als ge Halloween outfits of een Nieuwjaarsvideo moet bedenken of uw beste Afrikaanse dansmoves moet bovenhalen ☺! De laatste jaren zijn niet altijd even gemakkelijk geweest voor onze familie, maar ik ben er zo fier op dat we er net dan voor elkaar zijn! Special thanks ook aan mijn zussie en Hans, die helemaal vanuit Holland afgebroemd kwamen om mij op te peppen bij grote doctoraats-paniekaanvallen.

Mama en papa, woorden schieten tekort om jullie te bedanken, echt... Bedankt voor het warme nest waar ik nog steeds niet uit kan (en vooral niet wil) springen, voor jullie onvoorwaardelijke liefde en steun, voor het veilige gevoel dat jullie mij geven en het feit dat ik bij jullie altijd weer mijn batterijen kan opladen. Jullie zijn zonder twijfel de beste ouders die er bestaan! Ik kan alleen maar hopen dat ik het later even goed als jullie ga doen... Bedankt voor alle support tijdens mijn doctoraatsperiode (sorry voor alle vieze koekskes mam) en vooral om mij steeds weer te doen inzien dat het *maar* een doctoraat is en er zoveel belangrijkere dingen in het leven zijn (wat gaat ge zijn met uw gouden boekske als...). En paps, ik heb voortaan geen excuus meer om ook niet mee te komen luisteren naar uw ventilatie-, raam-, en/of verwarmingsuitweidsels (wat kiest ge: de lange grondbuis met de schimmels en bacteriën of het andere alternatief? You are the best ☺)

En dan last, maar definitely not least: Benneman, mijn grootste steun, mijn allerbeste vriend en ondertussen ook al 6 maanden wettelijk mijn ventje! Ik denk wel dat je weet dat ik dit zonder jou nooit of te nooit gekund had. Moest ik beginnen aan een opsomming van alle dingen die jij voor mij gedaan hebt in die vier doctoraatsjaren, dan zou ik nog een boekske van 300 bladzijden bij elkaar kunnen typen! Heel erg bedankt lieverd om zo goed voor mij te zorgen en mij elke dag weer het gevoel te geven dat ik speciaal ben. En zo hard sorry voor alle drama in de

vorige vier jaren (nu gaat ge mij heel hard moeten steunen want ik heb commissie, nu heb ik echt wel een moeilijke week want ik moet een coverletter schrijven, nu kan ik niet afwassen want ik zit in mijn laatste doctoraatsweken,....), voor het feit dat ik dacht dat het een goed idee was om Franse les, naailes, salsales en een ALO-opleiding te combineren met mijn doctoraat, voor mijn vele gesnauw in de stress pieken, voor de hopen rommel aan mijn kant van het bed, voor de koekjes die nooit de juiste dikte en grootte hadden, voor de bergen afwas die ik u alleen heb laten doen, voor het feit dat ge de laatste weken samengeleefd hebt met een vogelverschrikker in peignoir die vastgegroeid was aan haar laptop,... maar vooral omdat ik vaak niet genoeg besefte dat dit boekske niets is in vergelijking met alle grote stappen die wij de laatste vier jaar samen gezet hebben: samenwonen in de Rondestraat, een poezekop[1] adopteren, onze wettelijke trouw (na het mooiste en ontroerendste aanzoek ever...), het beginnen dromen en plannen van een eigen huisje in Testelt,... Er staan ons nog zoveel mooie dingen te wachten en ik kus elke dag mijn pollekes dat ik die samen met jou mag beleven. Ik zie je graag bollie!

[1] En kwestie van toch ook één voetnoot in mijn dankwoord te hebben: ook een bedankje aan Gonzie, mijn trouwste kompaan tijdens de laatste schrijfweken. Bedankt voor je dolle kattenfratsen die mij ook op stressmomenten deden lachen en de wijze raad dat slapen de beste hobby is die er bestaat!

Table of contents

List of tables

List of figures

1 General introduction

General background: The possible role of negative affective valence in (return of) fear

Epidemiological research shows that about 30 percent of the general population will develop an anxiety disorder at some point in life (Kessler, Koretz, Merikangas, & Wang, 2004). Not surprisingly then, a lot of research is invested in unveiling the processes that are responsible for the origin and maintenance of fear, as well as in the principles that constitute the core of successful treatment. The efficacy of exposure-based treatments for anxiety disorders is now irrefutable. Generally speaking, exposure treatment involves (gradually) confronting the patient with the fear-provoking object or situation until it no longer provokes feelings of anxiety. For some anxiety disorders like simple phobia, treatment success is so high that it is almost unequalled in our health services in general (e.g., Öst, 1989). Despite the general success of exposure-based treatments, however, some patients do not respond well to this type of treatment. Another remarkable observation is that a number of patients - in spite of apparently successful treatment - experience a return of symptoms (or even complete relapse) of fear and anxiety. Both observations constitute a challenge for clinical practice and fear research.

The present dissertation looks at fear from a learning theoretical perspective. Contemporary models of human classical conditioning, and more specifically of fear conditioning, provide a rich conceptual framework for the understanding of the etiology, maintenance and treatment of human fears and phobias (Craske, Hermans, & Vansteenwegen, 2006). The essence of these models is that they view classical conditioning as the acquisition of associations between memory representations. Research on classical conditioning originates from Pavlov's studies on salivation in dogs (Pavlov, 1927). Just before his dogs received food powder, Pavlov rang a bell. When the dogs received the food powder, they started to salivate. After several trials of ringing the bell before the presentation of the food powder, the dogs started to salivate upon hearing the bell, before the actual food powder was presented to them. The bell, a stimulus that was neutral to the dogs prior to training, is called the conditioned stimulus or CS. The food powder, which elicits a salivation reaction prior to the training, is called an unconditioned stimulus or US. The initial salivation reaction to the food powder is termed the unconditioned response or UR; the novel salivation response to the ringing of the bell due to its pairing with the food powder is called the conditioned response or CR. Contemporary conditioning models presume that during the conditioning trials, an association is formed between the memory representations of the CS and US.[2] As a result of this association, the CS acquires the ability to activate the representation of the US, which then results in the production of a CR. Pavlov's salivating dogs constitute an example of appetitive conditioning. As previously mentioned, the classical conditioning model can also shed light on how fear and phobias can be acquired. Consider the example of a child previously not scared of dogs, experiencing a dog bite. Like Pavlov's dogs, the child can form an association between the memory representation of the CS (the dog in this example) and the US (the painful dog bite and accompanying fear). Later confrontation with dogs will activate the child's representation of the painful event as well as the fear that is associated with it. From a conditioning perspective, exposure therapy can be viewed as the clinical analogue of extinction. In an extinction procedure, the CS is repeatedly presented without the US. As a result of those repeated CS-only presentations, conditioned responding will typically decline (Bouton, 1993; Pavlov, 1927). Return of fear after treatment then can be seen as the clinical analogue of return of conditioned responses after extinction.

In the context of the current dissertation, two recent insights from fear research are important. First, during the last two decennia it became clear that extinction does not reflect an 'unlearning' of the underlying CS-US association but rather involves the learning of a new association which temporarily suppresses the former one (Bouton, 2004). This original association can be re-activated by several 'post-extinction events', resulting in a return of the extinguished responses. Examples of such post-extinction events are the mere

[2] Note that not all learning psychologists would agree with the statement that conditioned changes in behavior are due to the formation of associations. Mitchell, De Houwer, and Lovibond (2009), for instance, have argued that Pavlovian conditioning might (also) result from the conscious formation and evaluation of propositions about CS-US relations.

passage of time ('spontaneous recovery'), a change of context ('renewal') or the presentation of US-only trials ('reinstatement') after extinction.

A second relevant insight is that during a fear acquisition procedure the meaning of the CS alters in two important ways. The CS not only becomes a valid predictor for the US (which results in its presentation being accompanied by increasing levels of fear), but also acquires a negative connotation through evaluative conditioning (e.g., Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2002). In the literature, these two types of learning are referred to as expectancy learning and evaluative learning. *Expectancy learning* is inferred when the presentation of the CS activates the expectation of US occurrence. For example, in a Pavlovian conditioning preparation in which a tone CS is repeatedly paired with a shock US, the tone may come to activate the expectation of immediate shock deliverance. Likewise, in the example of the child that was bitten by a dog, the sight of a dog may become established as a CS activating the expectation of being bitten again. The second type of learning, evaluative learning or *evaluative conditioning* (EC) refers to the observation that the mere contingent presentation of a neutral stimulus with a liked (disliked) stimulus, changes the valence of the originally neutral stimulus into a positive (negative) direction (for extensive reviews, see De Houwer, Baeyens, & Field, 2005; De Houwer, Thomas, & Baeyens, 2001). In the example of the tone-shock procedure, this implies that the initially neutral tone will acquire a negative connotation throughout conditioning. Likewise, a child that was bitten by a dog will no longer consider dogs as neutral but will experience them as negative.

Even though our examples clearly illustrate that expectancy learning and evaluative learning can both occur (and co-occur) in a conditioning procedure, these two types of learning seem to possess different functional characteristics (Baeyens & De Houwer, 1995). Most important for the present dissertation is the fact that several studies suggest that - compared to expectancy learning - evaluative learning is less susceptible or even resistant to extinction (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Hermans et al., 2002; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006). For example, using a differential fear conditioning paradigm in which one of two CSs (CS+) was consistently paired with a shock while the other one (CS-) was never paired with this US, Hermans et al. (2002) found that even though participants indicated after the extinction phase that they no longer expected the US to occur after the CS+, they still perceived this stimulus as negative.

If we translate this to a clinical situation, the findings concerning the weaker extinction susceptibility of evaluative learning suggest that exposure treatment for conditioned fear might successfully reduce the expectancy component of the fear (and therefore may lead to diminished fear reactions), but might leave the acquired affective meaning of the phobic object relatively unaltered. In our example of the child suffering from dog phobia, this would mean that after successful exposure therapy the child no longer expects to be bitten every time he/she passes a dog, but still somehow dislikes dogs. Clinical experience indicates that this differential outcome can indeed be observed (Marks, 1987).

Important for the current dissertation is the fact that this negative valence that outlasts extinction/exposure might not be without consequences. There is, for instance, clear evidence that negative stimuli are more easily associated with aversive outcomes than are neutral or positive stimuli (e.g., Hamm, Vaitl, & Lang, 1989). Research also suggests that negative valence is associated with action tendencies of escape and avoidance (e.g., Chen & Bargh, 1999). Given these findings, it is not unlikely that the remaining negative valence of a phobic object after exposure therapy might prompt avoidance behavior in patients, which in its turn might eventually initiate partial or full relapse. An alternative route through which residual negative associations could increase relapse risk follows from the emotion theory proposed by Lang (1995). According to Lang, all emotions can be situated in a two-dimensional space, with affective valence (positive/negative) and arousal (high/low) as crucial dimensions. Fear, for instance, is considered as an emotion that is characterized by a combination of negative valence and high levels of arousal. Following the finding that extinction/exposure leads to a significant decrease in arousal, but leaves negative stimulus valence intact (cf. Hermans et al., 2002), one could predict from the theory of Lang that fear may re-emerge relatively easily after extinction/exposure if the still negatively appraised fear object is encountered in an arousing context and negative valence and a higher level of arousal are thus recombined. Additionally, based on Lang's theory re-emergence of fear might even be expected if the experienced arousal is in fact unrelated to the US. In our example this would imply that if a child that was treated for dog phobia encounters a dog while he/she is aroused for reasons that are unrelated to the situation (e.g., a bad school report, too much sugar), this combination might reinstate the child's fear. Some recent experimental studies support the idea that the remaining negative valence of a fear stimulus after extinction/exposure might play a role in return of fear. Using a reinstatement paradigm in which unsignaled USs were presented after a differential fear acquisition and extinction procedure, both Dirikx et al. (2004) and Hermans et al. (2005) found the remaining negative valence of the reinforced CS (CS+) after extinction to be predictive for the return of fear for this stimulus after reinstatement. The more negative the CS+ remained after extinction, the more reinstatement was observed. Further evidence comes from a study in spider phobics by Huijding and de Jong (2005), who found residual negative associations after exposure treatment to be predictive for symptom return (i.e., return of overt avoidance behavior) at two-month follow-up.

So far, we have argued that the negative affective valence that a previously neutral stimulus acquired during a fear conditioning procedure might constitute a vulnerability factor for return of fear after extinction. In addition to impacting the more long-term outcome of an extinction or exposure treatment, the acquired negative valence of a fear object might, however, also influence the course of this treatment. It seems reasonable to assume that extinction learning in itself might also be hampered by the acquired negative valence of the fear object. However, this possibility has not been investigated yet.

The different ideas and findings concerning the possible impact of the negative valence that a stimulus acquires during fear conditioning are both theoretically and clinically relevant. Importantly, the above mentioned findings seem to suggest that patients with a fear disorder might benefit from a treatment that not only focuses on the disconfirmation of expectancies through exposure, but also targets the acquired negative valence of the fear object.[3] Removing the phobic object's negative connotation might enhance treatment progress, as well as lower relapse risk: the valence of the phobic object will no longer hamper exposure treatment and no longer constitute a vulnerability factor. In our example, this would imply that the short- and long-term outcome of exposure therapy might be improved if the therapist also works on reducing the child's negative evaluation of dogs.

Up until now, EC research has mainly focused on how (dis)likes can be *acquired* through conditioning. Far less research has been conducted on how conditioned preferences, once acquired, can be *altered*. Given the potential therapeutic[4] merit of techniques through which a person's existing (dis)likes can be changed, we believe it to be important that more research effort is invested in finding and studying such techniques. In the following section, we describe how the present dissertation has tried to fulfill this need.

---

[3] In this introduction, we focused on the possible co-occurrence of expectancy learning and evaluative learning in a fear conditioning procedure. Note that the presented ideas also hold for appetitive learning. From a clinical perspective, an appetitive learning procedure can provide more insight in the phenomenon of addiction as it makes clear how initially neutral stimuli (e.g., seeing injection needles, socializing with 'drinking buddies', the smell of chocolate) that are repeatedly paired with the administration of a drug (e.g., heroin, alcohol, chocolate) can become elicitors of conditioned craving. Also in an appetitive paradigm, CSs can acquire both signal value and a new affective connotation. Van Gucht, Baeyens, Vansteenwegen, Hermans, and Beckers (in press), for instance, applied a differential appetitive conditioning procedure in which one serving tray was followed by the eating of chocolate while another tray was not followed by this US. Results of this experiment indicated that participants learned to expect (and crave) the US after presentation of the CS+, but also that the CS+ acquired a positive valence throughout conditioning. Also here, the finding that evaluative learning is little susceptible to extinction implies that a cue exposure treatment might leave the acquired valence of the cue unaltered. In this case the remaining *positive* valence of the cue could function as a vulnerability factor for relapse (e.g., because positive stimuli are likely to evoke an approach tendency). Again, treatment outcome might be improved by taking into account the evaluative learning component in addiction.

[4] Note that appropriate methods for changing people's existing evaluations may also have their merit in the field of advertising, anti-discrimination programs, election campaigns, etc.

The present dissertation: Changing and assessing valence

In the current dissertation, two lines of research can be distinguished. In the following paragraphs, we will introduce these briefly. A more detailed overview of the relevant literature and the experimental studies that were conducted will be given further on, in the introductory chapters of each line (see the first chapters of Parts 2 and 3).

Line 1: Changing conditioned preferences through counterconditioning

In the previous section, we noted that in a fear conditioning procedure an initially neutral CS becomes a predictor for the aversive US but also acquires a negative affective valence. Furthermore, we saw that some recent ideas and findings indicate that this acquired negative valence might negatively impact the course and (short- or long-term) outcome of an extinction or exposure treatment. These findings then seem to suggest that treatment (outcome) could be improved by also targeting the evaluative learning component in conditioned fear. As previously noted, several studies suggest that evaluative learning is little susceptible or even resistant to extinction (e.g., Diaz, Ruiz, & Baeyens, 2005; Hermans et al., 2002; Vansteenwegen et al., 2006). It appears to be the case that once a stimulus has acquired a valence as the result of being paired with a liked or disliked stimulus, this acquired valence cannot readily be changed by repeatedly presenting the stimulus on its own. A technique that might be more successful in altering the CS's conditioned valence is counterconditioning (CC). In a CC procedure, a CS is not presented without US, but is paired with a new oppositely valenced US. In our dog phobia example, a CC procedure could for instance entail presenting the child with his/her favorite type of candy every time he/she encounters a dog.

A review of the EC literature (see Chapter 1, Part 2) revealed that even though CC seems to represent a promising approach (e.g., Baeyens, Eelen, Van den Bergh, & Crombez, 1989), the evidence for the effectiveness of this technique is rather scarce and preliminary. Therefore, the primary aim of the experimental studies conducted in the first (and main) line of research was to further explore the effect and properties (e.g., time- and context-sensitivity) of evaluative CC.

As our review indicated that the current work and knowledge on evaluative CC is rather limited, we decided to focus on the topic of CC in itself before studying its impact on the course or outcome of an extinction or exposure treatment. The little available empirical research also prompted us to first investigate the effects and properties of a CC treatment on *newly acquired moderately valenced* conditioned (dis)likes before studying its impact on *long-existing* or *deeply routed* (dis)likes like the negative valence of a fear object.[5]

Five experimental studies were conducted in this line. These are presented in Part 2 of this dissertation.

Line 2: Assessing valence indirectly and online

As made clear in the previous section, the main focus of the current dissertation was on evaluative learning, and more specifically on how conditioned (dis)likes are (formed and) changed. Obtaining a good measurement of people's evaluations is crucial if one wants to study evaluative learning processes. Against this background, in a second line of research, we conducted a number of studies on the (indirect and online) assessment of people's evaluations.

The most popular valence measures in EC studies are self-report ratings and indirect reaction time (RT) tasks. In EC studies, these measures are typically administered in a session that precedes and/or follows the conditioning phase (i.e., as pre-post measures). In Part 3, we discuss the disadvantages of such pre-test post-test designs and argue that it might be preferable (or at least interesting) to assess valence *during* conditioning, rather than only before and after. Since the currently existing behavioral valence measures do not lend themselves well to be integrated in an ongoing conditioning procedure, we focused in our second line of research on the development of an indirect evaluative RT task that *can* be administered online.

[5] From the outset, we were aware of the fact that our choice for a 'bottom-up' approach in which we gradually advance from more basic to more complex research questions would imply that most likely not all the research questions that were introduced in the previous section could be addressed in the limited time span of the doctoral project. We acknowledge that the current dissertation represents only a first (but according to us important) step within the broader research goal of understanding the role of stimulus valence in the origin and treatment of fear.

Three experimental studies were conducted in this line. These are discussed in the third part of this dissertation.

References

Baeyens, F. & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical conditioning: A reply to Davey (1994). *Behaviour Research and Therapy, 33,* 825-831.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective-evaluative value: Conservative but not unchangeable. *Behaviour Research and Therapy, 27,* 279-287.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114*, 80-99.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*, 485-494.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25,* 215-224.

Craske, M. G., Hermans, D., & Vansteenwegen, D. (2006), *Fear and learning: Basic science to clinical application.* Washington, DC: American Psychological Association.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*, 161-174.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127,* 853-869.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*, 245-268.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11,* 549-554.

Hamm, A. O., Vaitl, D., & Lang, P. J. (1989). Fear conditioning, meaning, and belongingness - A selective association analysis. *Journal of Abnormal Psychology, 98,* 395-406.

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S. P. Shohov (Ed.), *Advances in Psychology Research* (pp.17-41). New York: Nova Science.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*, 533-551.

Huijding, J., & de Jong, P. J. (2005). Changes of automatic and self-reported spider phobia-related affective associations following treatment and the prediction of symptom return. Unpublished doctoral dissertation, Rijksuniversiteit Groningen.

Kessler, R. C., Koretz, D., Merikangas, K. R., & Wang, P. S. (2004). The epidemiology of adult mental disorders. In B. L. Levin, J. Petrilia & K. D. Hennessy (Eds.), *Mental health services: A public health perspective, Second Edition.* New York: Oxford University Press.

Lang, P. J. (1995). The emotion probe - studies of motivation and attention. *American Psychologist, 50,* 372-385.

Marks, I. M. (1987). Fears, phobias, and rituals: Panic, anxiety and their disorders. Oxford: Oxford University Press.

Mitchell, C. J., De Houwer, J., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*, 183-198.

Öst, L. G. (1989). One-session treatment for specific phobias. *Behaviour Research and Therapy, 27,* 1-7.

Pavlov, I. P. (1927). *Conditioned reflexes.* Oxford, UK: Oxford University Press.

Van Gucht, D., Baeyens, F., Vansteenwegen, D., Hermans, D., & Beckers, T. (in press). Counterconditioning reduces cue-induced craving and actual cue-elicited consumption. *Emotion.*

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 71-79.

Changing evaluations through counterconditioning

2

The main aim of the experimental studies conducted in the first line of research was to further explore the effect and properties of counterconditioning (CC) in evaluative conditioning (EC). Five experimental studies will be discussed in this part of the dissertation. All studies, except for the first and fourth one, are formatted as manuscripts that stand on themselves. Inevitably, this entails a certain degree of redundancy in the ensemble of the dissertation. Still, effort has been made to indicate cross-links between chapters.

The first chapter provides an overview of the literature on the acquisition and change of (dis)likes from a conditioning perspective. Although some (dis)likes are genetically determined, most of our preferences stem from learning. In this chapter, we discuss how people can acquire new (dis)likes through associative learning. From an applied point of view, it is not only interesting to understand how people acquire new preferences, but also to know how existing preferences can be influenced and altered. Three possible procedures to change conditioned valence have been identified in the EC literature: extinction, US-revaluation and CC. A remarkable finding is that, even though EC resembles Pavlovian conditioning procedurally, it differs in being less susceptible to extinction. A review of the literature on US-revaluation and CC suggests that both procedures provide promising approaches for changing previously acquired conditioned valence. Both techniques are, however, understudied. The present dissertation aimed at a deeper investigation of the effect and properties of CC in EC.

If one wants to investigate whether conditioned valence can be altered with a CC procedure, having a stable conditioning paradigm in which participants can learn new (dis)likes is a prerequisite. The main goal of the experiments presented in Chapter 2 was to find such a robust EC paradigm. Three different EC procedures were tested. In Experiment 1, a picture-picture procedure was put to the test in which both the CSs and USs were pictures of human faces. In a second experiment, we tried out two gustatory EC procedures. Both studies had a similar design and consisted of three consecutive phases: (1) a pre-measurement of the valence of the CSs, (2) an EC phase in which two CSs were consistently paired with a positive US while two other CSs were always combined with a negative US, and (3) a post-measurement of CS valence. Valence was measured directly with evaluative ratings and indirectly with an affective priming task (Fazio, Sanbonmatsu, Powell, & Kardes, 1986). In both experiments we expected the CSs to acquire the valence of the US with which they were paired during the conditioning phase.

In Chapter 3, I present a study (Experiment 3) in which we investigated whether conditioned preferences can be altered with a CC procedure. Based on the results of the previous two studies, we chose to work with a picture-flavor paradigm. In Experiment 3, we compared the effect of further conditioning, extinction and CC on recently acquired conditioned preferences. We expected that only the CC treatment would be effective in altering the previously acquired valence of the CSs. Valence was again measured using rating scales and a priming task. In this experiment, we also explored whether counterconditioned preferences remain intact over time. A frequently cited and fascinating property of EC effects is their stability. Baeyens, Crombez, Van den Bergh, and Eelen (1988), for example, found that conditioned evaluative responses remained intact two months after initial conditioning. An interesting question is whether the same applies for counterconditioned (dis)likes: will they persist or regress to their initial meaning of time? If the latter is the case, this would suggest that the possible beneficial therapeutic effects of a CC procedure on (return of) fear might also dissipate over time.

In Chapter 4, I present a study (Experiment 4) in which we explored whether evaluative learning (and more specifically CC) is sensitive to context manipulations. Several studies on ABA renewal have demonstrated that this is the case for expectancy learning (e.g., Vansteenwegen et al., 2005). In these experiments, participants learn certain contingencies in a first context (referred to as the A context), learn different contingencies in a second context (referred to as the B context) and then are tested for their expectancies upon return to the initial acquisition context (A). Expectancy indices typically show that participants expect the original contingencies to be valid again in the acquisition context. Results of ABA renewal experiments make clear that people learn context-specific *US expectancies* in such a design. In Experiment 4 we wanted to investigate whether people also acquire context-specific *evaluations* when CC takes place in a different context than the one in which the initial conditioned preferences were acquired. If the effect of the CC treatment would be lost upon return to the original acquisition context, this would be an indication that the possible therapeutic effects of this type of treatment might also be context-specific (e.g., specific to the therapeutic context) and thus only temporary.

Finally, in Experiment 5 (presented in Chapter 5), we went one step further and examined the impact of a CC procedure on return of fear using a reinstatement procedure in mice. Not only in humans, but also in animals little is known about the precise role of stimulus valence in fear conditioning. Similar to studies in humans, dissociations have been found between expectancy and evaluative learning (e.g., Campbell, Capaldi, Sheffer, & Bradford, 1988). Therefore, it seemed interesting to also test our hypothesis that CC trials might help to reduce return of fear, in mice. In previous research in our lab (Dirikx et al., 2007), return of fear due to a reinstatement procedure (i.e., unsignaled US presentations after extinction) was observed in mice using a differential conditioned suppression paradigm. In a first phase, one CS (CS+) was consistently paired with a foot shock US, while another CS (CS-) was not, resulting in selective suppression of previously trained instrumental behavior during the CS+. After the extinction phase, half of the animals (reinstatement group) received unsignaled USs while the other half (control group) did not. A differential return of conditioned responding was observed in the reinstatement group only. In Experiment 5, we used the same differential conditioned suppression paradigm, but added an extra phase between the extinction and reinstatement phase. In this extra phase, half of the mice received CC trials during which the CS+ was paired with food pellets, while the other half received further extinction trials. We expected to find less return of fear for the CS+ after reinstatement in the group of mice that received CC trials as compared to the group that underwent further extinction.

Each chapter includes a discussion section. The results of all studies will, however, also be recapitulated and discussed in the General Discussion of this dissertation (Part 4).

References

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact: Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10,* 179-199.

Campbell, D. H., Capaldi, E. D., Sheffer, J. D., & Bradford, J. P. (1988). An examination of the relationship between expectancy learning and preference conditioning. *Learning and Motivation, 19,* 162-182.

Dirikx, T., Beckers, T., Muyls, C., Eelen, P., Vansteenwegen, D., Hermans, D., & D'Hooge, R. (2007). Differential acquisition, extinction and reinstatement of conditioned suppression in mice. *The Quarterly Journal of Experimental Psychology, 60,* 1313-1320.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229-238.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behaviour Research and Therapy, 43*(3), 323-336.

The acquisition and change of (dis)likes from a conditioning perspective

For all living organisms, it is vital that they are able to respond in a different way to objects and events that are good for them than to those that are bad for them. As all other organisms, humans therefore constantly evaluate objects as being good or bad, liked or disliked. Evaluation of stimulus valence is so fundamental that Lang, Bradley, and Cuthbert (1990) have proposed that the brain uses it as a basic category for organizing information. Because virtually all objects or events can be endowed with a certain valence (Osgood, Suci, & Tannenbaum, 1957), preferences influence many aspects of our life. They impact upon the foods we eat, the products we buy, the stimuli we approach or avoid and the people with whom we spend time. Our emotions, too, are tightly linked to our preferences: they generally come about only when certain valenced objects or events are involved (e.g., Sherer, 1993). Given this pervasive impact, it is no wonder, then, that the study of likes and dislikes has attracted the interest of scholars in many of psychology's subdisciplines (e.g., learning psychology, Martin & Levey, 1978; social psychology, Walther, Nagengast, & Trasselli, 2005; consumer science, Gibson, 2008; clinical psychology, Hermans, 1998).

In order to understand, predict, and influence human behavior, it is important to know how preferences are formed and can be changed. Although some preferences are genetically determined, most stem from learning that took place during the lifetime of the individual (Rozin, 1982). In the present dissertation, we focus on one particular pathway through which evaluations can be acquired and changed: associative learning. Before we take a closer look at this pathway, we clarify how the term 'evaluation' is used in this thesis.

Evaluation operationally defined as an effect

As mentioned above, evaluation has been studied in many areas of psychology. In these different domains, a variety of terms and definitions have been used to describe the process or effects of evaluation, including concepts like 'attitude', 'affect', 'valence', '(dis)likes', etc. In the present paper we equate all these different concepts with the term 'evaluation' and define this last concept following De Houwer (2009), who argued that evaluation can be <u>defined</u> either as the *process* of determining the liking of an object or as *responding* in an evaluative manner to objects. <u>Studying</u> evaluation, however, always boils down to research on evaluative responding, because the presence of an evaluation process can only be inferred on the basis of the presence of an evaluative response. For this reason, we follow De Houwer in defining evaluation as an effect, that is, as the occurrence of an evaluative response that is caused by the presence of an object, rather than as a mental process. Possible evaluative responses are, for instance, selecting a number on a Likert scale that represents one's liking of an object (e.g., Likert, 1932), physiological responses such as changes in the activity of facial muscles (e.g., Lang, Greenwald, Bradley, & Hamm, 1993) or behavioral responses such as the modulation of the speed with which a target is categorized as positive or negative in an affective priming task (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986). In principle, any response can be an evaluative response as long as there are arguments to back up the claim that the response is determined by the liking of objects (e.g., empirical evidence that shows that the response changes as the result of (manipulations that lead to) changes in the liking of objects). In Part 3 of this dissertation, we give an overview of the most commonly used evaluative responses in experimental research. Note that backing up the claim that a response is an evaluative one is not always evident. As we will see in Part 3, different opinions exist, for instance, on whether the startle reflex represents an evaluative response or not. In their 1990 study, Lang et al. observed affective modulation of the startle reflex with startle blink facilitation for startle probes presented during unpleasant pictures and startle blink inhibition for probes presented during pleasant pictures. Based on these data, several authors used the startle response as an index of stimulus valence (e.g., Purkis & Lipp, 2001). Cuthbert, Bradley, and Lang (1996), however, observed that modulation was only found for pictures with high levels of judged arousal, which prompted some authors (e.g., Vansteenwegen, Crombez, Baeyens, & Eelen, 1998) to conclude that startle modulation is not a good index of valence - at least not when stimulus arousal is low.

An important advantage of defining evaluation as a response is that this definition remains neutral about the precise nature of the processes that underlie the generation or change of evaluative responses. Therefore, the validity of this definition does not depend on the validity of our current understanding of the psychological processes and representations underlying evaluation and evaluation change (a topic of considerable debate as will become clear throughout this dissertation).

The origin of evaluations: Evaluative responding can be learned

As stated in the introductory paragraph, people constantly evaluate their environment. We interact with a world loaded with valence. It is difficult to imagine a person who would be impartial towards all that he or she encounters. A crucial question that arises then is: Where do these evaluative responses come from? What is the origin of the likeable or dislikeable nature of an event?

Empirical evidence suggests that for an infant, the world is quite undifferentiated. Only few stimuli are *innately* perceived as clearly positive or negative in valence (e.g., pain, heat, sweet or bitter tastes; Rozin & Millman, 1987). The vast majority of our likes and dislikes are *acquired* during life. This is well demonstrated by the fact that people can grow to like stimuli that evoke a universal negative response in babies (e.g., chili peppers) and by the large cultural differences in (dis)likes (see Rozin, 1982, for a review).

Although there are several pathways through which valence can be acquired, a principal route is associative in nature, and is based on the contiguous or contingent pairing of originally neutral stimuli with events that already have a positive or negative valence. For instance, children may learn to like orange flavor because of the mere repeated co-occurrence of orange flavor with the agreeable sweetness of sugar, abundantly present in the average orange soft drink. In the literature the term ‘Evaluative Conditioning’ (EC) is used to refer to changes in the evaluation of a stimulus that result from pairings with other positive or negative stimuli (De Houwer, Thomas, & Baeyens, 2001).

In the case of EC, the acquired evaluative response is based on the co-occurrence or pairing of two stimuli. As mentioned above, other pathways exist through which humans learn evaluative responses. For instance, research on the mere exposure effect has demonstrated that repeated, unreinforced exposure to a stimulus can result in increased liking for that stimulus (for recent theoretical perspectives on the phenomenon, see Winkielman, Schwarz, Fazendeiro, & Reber, 2003; Zajonc, 2001). Further, several studies suggest that a stimulus can also acquire a certain valence through its pairing with positively or negatively valenced (consequences of) actions. Fazio, Eiser, and Shook (2004), for example, showed that stimuli that signal that an action will have a negative outcome are liked less then stimuli that signal that an action will have a positive outcome. Also the valence of the actions themselves can influence evaluative responding to stimuli associated with those actions. For instance, repeatedly approaching (avoiding) a neutral stimulus seems sufficient to change its evaluation into a positive (negative) one (Cacioppo, Priester, & Berntson, 1993; Woud, Becker, & Rinck, 2008; also see Bem, 1972). The so far discussed routes for acquiring an evaluative response towards a stimulus are all based on *direct experience* with the stimulus under concern (Fazio & Zanna, 1981). A (change in the) evaluative response towards a certain object can, however, also result from *indirect experience* with this object, that is, from information that one receives from others via verbal instruction (like in persuasion research, see Crano & Prislin, 2006 for a review) or that one picks up through observation (e.g., Baeyens, Vansteenwegen, De Houwer, & Crombez, 1996).[6]

The present dissertation focuses on the associative pathway of preference acquisition and change. Therefore, in the next section, we will take a closer look at the research that has been conducted on EC.

A closer look at evaluative conditioning

As was mentioned, EC refers to changes in the evaluation of a stimulus that result from pairings with other positive or negative stimuli. In a prototypical EC experiment, a neutral stimulus (generally referred to as the conditioned stimulus or CS) is presented repeatedly with a subjectively liked or disliked stimulus (the unconditioned stimulus or US). After repeated pairings, the CS valence typically shifts in the direction of the US valence. At a procedural level, EC can be regarded as a form of Pavlovian conditioning in that it involves a change in the response to the CS that results from the pairing of this stimulus with a US. Whereas Pavlovian conditioning can refer to a change in *any* type of response,

[6] The here-presented classification of pathways through which valence can be acquired is based on De Houwer, 2009.

EC concerns only a change in the evaluative responses to the CS, that is, a change in the *liking* of the CS (De Houwer, in press).

The first demonstrations of EC effects date back more than 50 years (Razran, 1954; Staats & Staats, 1957). Razran, for instance, repeatedly exposed participants to various political slogans under one of three conditions: (a) eating a free lunch, (b) inhaling unpleasant smells, or (c) sitting in a neutral setting. Both before and after this exposure, participants' evaluation of each slogan was assessed. Razran found that the slogans that were paired with the free lunch were rated more positively than the slogans that were associated with the aversive odors, while no change in valence was observed for the slogans that were paired with the neutral setting. Modern EC research was strongly inspired by the work of Levey and Martin (1975). They introduced the so-called picture-picture paradigm that is still frequently used today. In the study of Levey and Martin, participants were first required to sort a set of postcard pictures into a liked, disliked, and neutral pile. In a subsequent acquisition phase, initially neutral postcards (CSs) were presented together with either liked, disliked or other neutral postcards (USs). Subsequent liking ratings showed that the valence of the CSs that were paired with a liked or disliked US had changed in the respective direction of the US valence.

Since these early demonstrations, preference learning through EC has been examined in a larger number of studies in different areas (De Houwer et al., 2001). Despite the huge body of EC research, however, our understanding of the phenomenon is still very limited (De Houwer, 2007). In recent years, general agreement has emerged that EC is a genuine phenomenon (De Houwer, Baeyens, & Field, 2005). EC effects have been obtained using a variety of stimuli and paradigms (for reviews see De Houwer et al., 2001; 2005). At the same time, however, failures to observe EC effects have been haunting the field (e.g., Field & Davey, 1999; Rozin, Wrzesniewski, & Byrnes, 1998; also see Experiment 1 in Part 2 of this thesis). Identifying boundary conditions and developing robust paradigms for studying EC therefore form two important pathways in current research. A lot of research has also been devoted to examining the functional properties of EC. Initial research suggested that, in comparison to other forms of Pavlovian associative learning (e.g., preparatory conditioning), EC appeared to have several unusual functional characteristics. Some authors, for instance, found EC to not depend on whether the participants were aware of the CS-US pairings (contingency awareness, see for instance Baeyens, Eelen, & Van den Bergh, 1990; Dickinson & Brown, 2007; Walther & Nagengast, 2006). Several studies also found EC to be resistant to extinction (unreinforced presentations of the CS alone after acquisition, see for instance Diaz, Ruiz, & Baeyens, 2005; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006) and insensitive to modulation and cue competition (e.g., Baeyens, Crombez, De Houwer, & Eelen, 1996; Baeyens, Hendrickx, Crombez, & Hermans, 1998; Beckers, de Vicq, & Baeyens, 2009). Recent studies, however, suggest that EC - at least sometimes - *does* show the same functional properties as other forms of Pavlovian learning. Pleyers, Corneille, Luminet, and Yzerbyt (2007) and Stahl, Unkelbach, and Corneille (2009), for instance, found EC effects only for CSs for which participants could report the associated US. Lipp and colleagues (Hardwick & Lipp, 2000; Lipp, Neumann, & Mason, 2001; Lipp, Oughton, & Lelievre, 2003), on the other hand, found evidence for extinction, modulation and blocking effects in EC.

Hitherto, it is not really clear how to reconcile the wide range of seemingly inconsistent findings in the EC literature. Some authors have suggested that methodological or procedural factors might contribute to the discrepant findings (e.g., Lipp et al., 2003). EC studies indeed differ on nearly all possible *procedural* and *methodological* parameters (e.g., number of CS-US pairings, dependent measures, duration of the CS and the US, modality of the stimuli, instructions, etc.). Other authors have argued that the opposing results might reflect the operation of different mental *processes* (De Houwer, 2007; De Houwer et al., 2005). Several processes have been postulated to underlie EC effects. Some theories (e.g., the referential account of Baeyens, Eelen, Crombez, & Van den Bergh, 1992) assume that EC depends on the automatic formation of associations in memory between elements of the CS and US representation. Other accounts, however, emphasize the role of higher-order mental processes, like propositional reasoning, in EC effects (e.g., De Houwer, 2007, 2009). A possible explanation for the discrepant findings in the EC literature might then be that EC effects can come about in multiple ways and that different mental processes were at play in studies that found opposing results. Moreover, which processes are involved might depend on procedural parameters like the number of pairings, type of stimuli, instructions, etc.

A detailed description of all the findings and theories in EC research is beyond the scope of this introductory chapter. In the next section, we will focus on how conditioned preferences, once acquired, can be changed. Against that background, we will take a closer look at the experimental research on extinction, US-revaluation and CC in EC. Other aspects of the EC debate will, however, be covered in the experimental chapters (Chapters 2-5) and the general discussion section (Part 4). For detailed reviews and a meta-analysis of EC research, we refer to De Houwer et al. (2001, 2005) and Hofmann, De Houwer, Perugini, Baeyens, and Crombez (in press).

Changing conditioned preferences

In the previous section, it became clear that three main questions have guided EC research (Hofmann et al., in press). First, a majority of the studies examined whether EC is an authentic and general phenomenon. Second, researchers have focused on the functional properties of EC, mainly in the quest to determine whether EC is a unique form of Pavlovian conditioning. A third, and more recent question concerns what processes are involved in EC.

From a more applied point of view, an interesting question is how (conditioned) preferences, once acquired, can be influenced and altered.[7] Appropriate methods for changing people's existing evaluations are of interest to many scientists and practitioners due to their wide-ranging applicability in fields such as advertising, anti-discrimination programs, or election campaigns, to name just a few. In the general introductory chapter, we hypothesized that these methods might also have their use in clinical practice, for instance in the treatment of (return of) fear.

Three possible procedures to alter conditioned valence have been identified in the EC literature: extinction, US-revaluation and CC.

[7] Note that this question is not independent of the previous ones. Knowledge on the properties of and processes behind EC can provide valuable suggestions on what strategies might (not) work to alter conditioned valence.

Extinction

Studies on *Pavlovian* conditioning typically show that a conditioned change in behavior can be reversed[8] by presenting the CS on its own after the CS-US trials (e.g., Hamm & Vaitl, 1996). This phenomenon is known as extinction and forms the basis of exposure therapy for phobias and anxiety disorders (Craske, Hermans, & Vansteenwegen, 2006).

As was mentioned above, surprisingly, several studies found the magnitude of *EC* effects to be unaffected by extinction training. For instance, in the context of the standard picture-picture paradigm, Baeyens and colleagues (Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Baeyens, Eelen, Van den Bergh, & Crombez, 1989) found that 5 and even 10 unreinforced CS-only presentations did not have any impact on the evaluative value that these CSs acquired as the result of 10 previous CS-US pairings. Caution is required, however, when interpreting the findings of these early studies as some authors have argued that these suffered from some important methodological flaws (e.g., Field & Davey, 1999). In these studies, for example, CS-US pairs were not randomized but arranged by the experimenter on the basis of perceptual similarity. Field and Davey (1999) demonstrated that such a CS-US assignment can result in artifactual EC effects. They obtained conditioning-like changes in the evaluation of the CSs when the CS-US pairs were constructed like this even in those participants who were never exposed to actual CS-US pairings. Therefore, the observed changes in valence in the studies of Baeyens and colleagues might have been due to similarity effects rather than to associative learning. If changes in liking are not based on associative learning, then it is not surprising that a removal of the association has no effect on these acquired (dis)likes.

The results of Baeyens et al. (1988, 1989) have, however, been confirmed in more recent and methodological rigorous studies using a variety of conditioning preparations (e.g., a picture-picture paradigm, De Houwer,

---

[8] At least in performance. As was mentioned in the general introductory chapter, recent Pavlovian conditioning research suggests that extinction does not reflect unlearning, but rather involves new learning. Phenomena like rapid reacquisition, spontaneous recovery, renewal and reinstatement indicate that the initial conditioned response can reappear under certain circumstances (Bouton, 2004). Also see further in this chapter.

Baeyens, Vansteenwegen, & Eelen, 2000; Dwyer, Jarratt, & Dick, 2007; Field, 2006; a flavor-flavor paradigm, Baeyens, Crombez, Hendrickx, & Eelen, 1995; a picture-shock paradigm, Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004). The main dependent variable in all these studies were self-reported evaluative ratings. Therefore, demand effects cannot be excluded. Several studies, however, also found no impact of an extinction procedure on EC when an indirect measure (i.e., an affective priming task) was used (e.g., Diaz et al., 2005; Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2002; Vansteenwegen et al., 2006). In the latter two studies, evidence for extinction-resistant evaluative learning was even obtained when measures more typical of Pavlovian preparatory conditioning (i.e., US-expectancy ratings and skin conductance) did show complete extinction.

Interestingly, Lipp and colleagues (Lipp et al., 2003; Lipp & Purkis, 2006; also see Blechert, Michael, Vriends, Margraf, & Wilhelm, 2007) found that when evaluative ratings were collected *during* the extinction phase, extinction effects *did* occur. In all the studies reporting extinction-resistant evaluative learning, valence was assessed *after* (rather than during) the extinction phase. According to Lipp et al. (2003), the post-test in these studies might have reflected a return of conditioned responding (i.e., renewal) caused by the context shift from the ongoing conditioning paradigm to the post-rating context rather than extinction resistance. In a more recent study, however, Blechert, Michael, Williams, Purkis, and Wilhelm (2008) failed to find extinction of EC even when liking of the CSs was measured during the extinction phase. As we will see in Part 3 (Experiment 2), also in our own research we found no evidence for extinction of evaluative learning when valence was assessed online (i.e., during conditioning) with an affective priming task.

A recent meta-analysis on EC (Hofmann et al., in press) also addressed the extinction issue and revealed that, across studies, substantial EC effects could still be observed after extinction training. Fine-grained analyses did, however, show that CS-only trials reduced the magnitude of EC effects. According to Hofmann et al. (in press), these findings suggest that even though EC might not be ‘resistant’ to extinction in the true sense of the word, extinction might occur at a slower rate in EC than in other forms of Pavlovian conditioning.

In sum, although the evidence is somewhat mixed, several studies suggest that the evaluative value that a stimulus acquired through EC might not (always) be easily changed through an extinction procedure. From this perspective, it seems worthwhile to look for alternative procedures to change conditioned valence. US-revaluation and CC might constitute such alternatives.

US-revaluation

In a US-revaluation procedure, a US's valence is changed in the absence of the CS after conditioning has been completed. Depending on whether the value of the US is increased or decreased, this procedure is also called US-inflation or US-deflation. In *Pavlovian* conditioning procedures, post-conditional changes in the valence of the US have been found to alter the conditioned response to the CS as well, even though the CS itself was never paired with the revaluated US (e.g., Delamater & LoLordo, 1991).

Baeyens, Eelen, Van den Bergh, and Crombez (1992), and more recently Walther, Gawronski, Blank, and Langer (2009) demonstrated that *EC* is also sensitive to US-revaluation. Walther et al., for instance, employed a picture-picture paradigm in which neutral faces (CSs) were repeatedly paired with either positive or negative faces (USs). Subsequently, the valence of the US faces was altered by presenting the positive faces with negative information and the negative faces with positive information. Results showed that the revaluation of the USs not only resulted in a reversal in US valence, but also led to corresponding changes in the valence of the pre-associated CSs. This revaluation effect was evidenced by rating data and affective priming data.

The findings of Hammerl, Bloch, and Silverthorne (1997) can also be seen as evidence for US-revaluation effects in EC. Using pictures of outdoor fountains and sculptures as CSs and USs, Hammerl et al. found that US-only presentations after the acquisition phase reduced the magnitude of EC effects. According to the authors, the USs lost some of their affective value due to the repeated presentations after acquisition (i.e., the USs were deflated). Like in the studies of Baeyens et al. (1992) and Walther et al. (2009) this US-revaluation experience subsequently affected the liking of the pre-associated CSs.

Note, however, that Baeyens, Vanhouche, Crombez, and Eelen (1998) failed to obtain US-revaluation effects in evaluative flavor-flavor conditioning. In two experiments, Baeyens et al. (1998) paired one of two flavors with an unpleasant taste (Tween20; a bitter soapy tasting chemical), while the other one was never paired with this US. After the acquisition phase, participants were exposed to a series of inflated US trials, consisting of a very strong concentration of Tween20 presented without the CS flavors. Even though participants experienced this higher dose of Tween20 as more aversive than the original acquisition US, the magnitude of the observed EC effects after US-inflation was equal to that of a control group for who the US was not revaluated. Like for other conflicting results in the EC literature, it is not really clear why Baeyens et al.'s (1998) results differ from the findings that were obtained with the picture-picture paradigm. More research is needed to clarify these discrepant findings.

Counterconditioning

In studies on US-revaluation, the identity of the US with which a CS is paired is kept stable whereas the valence of the US is changed. In studies on CC, both US identity and US valence are changed. More particularly, in such a procedure, the CS is paired with a US having a valence opposite from the original US. Several *Pavlovian* conditioning studies provide evidence that in a CC procedure, the originally learned conditioned response to the CS can become replaced by a response appropriate to the second US (e.g., Bouton & Peck, 1992; Peck & Bouton, 1990).

Baeyens et al. (1989) were the first to investigate the impact of CC on *evaluative learning*. Using a picture-picture paradigm, these authors observed that the conditioned positive/negative valence of a CS could be eliminated or even reversed by pairing it with a new negative/positive US. These results should be interpreted with caution, however, as also in this study, CS-US assignment was not random but based on perceptual similarity. Furthermore, Stevenson, Boakes, and Wilson (2000) failed to replicate the results of Baeyens et al. (1989) using an odor-taste paradigm. In their study, Stevenson et al. (2000) observed no difference in liking between an odor mixed with citric acid (i.e., an aversively sour tastant) only and an odor first mixed with citric acid and subsequently with sucrose. This replication failure might, however, relate to their choice of US. Possibly their intended positive US (i.e., sucrose) was not perceived as such by all participants as the authors also observed no increased liking for an odor consistently paired with sucrose.

The findings of Baeyens, Hendrickx et al. (1998) can also be interpreted as evidence for CC effects in EC. This experiment was aimed at investigating whether EC is sensitive to feature positive occasion setting. In a feature positive schedule, a 'target' stimulus A is reinforced if, and only if, it is accompanied by another stimulus (the 'feature' X). In Pavlovian conditioning, this reinforcement schedule is known to result in a conditioned response if the CS is preceded by X but not when the CS is presented alone (e.g., Holland, 1983, 1991; Rescorla, 1985, 1991). In several experiments, Baeyens and colleagues failed to obtain similar results in a number of flavor-flavor EC studies. In one of their experiments (Experiment 4), participants were exposed to a target flavor together with the aversive tastant Tween20 when it was preceded by the feature flavor X but received the target flavor together with sugar when it was preceded by plain water. For those participants who liked sugar, no valence shifts were observed, whereas participants who disliked sugar developed an unmodulated dislike for the target flavor. Thus, when sugar was liked it counteracted the effect of Tween20 because the target flavor was paired equally often with a liked (i.e., sugar) and disliked (i.e., Tween20) US. However, when sugar was disliked, it supported the effect of Tween20 because the target flavor was in effect always paired with a negative US (i.e., either sugar or Tween20).

Finally, Lipp and Purkis (2006) found a CC treatment to be effective in altering previously acquired conditioned valence in an experiment using online valence ratings as dependent measure. In this study, one of two geometrical shapes was paired with a happy face while the other shape was paired with an angry face. After 40 experimental trials, the contingencies switched. Results indicated that participants' initially acquired evaluations changed (more precisely: reversed) after the contingency switch, be it at a slower rate than during the first 40 acquisition trials. Note that in all the above discussed studies, the main dependent variable were self-report evaluative ratings. Hence, demand effects might have been at play.

Apart from the just-mentioned studies, surprisingly little research has been conducted on CC as a strategy for changing conditioned preferences. Recently, however, some studies have focused on (partially) related matters, like the impact of evaluative CC on *long existing* attitudes (e.g., self-esteem, Baccus, Baldwin, & Packer, 2004; Dijksterhuis, 2004; the negative attitude towards spiders in spider phobics, Eifert, Craill, Carey, & O'Connor, 1988; ageism, Karpinski & Hilton, 2001; racial prejudice, Olson & Fazio, 2006). Olson and Fazio (2006), for example, found that repeated pairings of black faces with positive stimuli and white faces with negative stimuli reduced racial prejudice to black people as measured by the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998; see also Part 3 of this dissertation). In all these studies, it is unclear, however, how and when participants acquired their 'a-priori' preferences. Moreover, participants might already have had evaluatively heterogeneous experiences with the attitude objects that were targeted in those studies (e.g., it is likely that people have both positive and negative associations with members of another racial group). Therefore, these reports provide little information on whether CC operates similarly for (unambivalent) preferences that were (recently) acquired through EC.

Also, some social psychology studies have investigated the impact of counter-attitudinal information on recently acquired attitudes using an *impression formation paradigm*. In this procedure, participants received information about whether certain positive and negative behaviors were (un)characteristic of a fictional person. In several experiments, Rydell and colleagues (Rydell & Gawronski, 2009; Rydell & McConnell, 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007; also see Kerpelman & Himmelfarb, 1971) observed that participants' experimentally induced attitudes could be changed with additional contradicting information. Nevertheless, it remains unclear whether a (more typical) EC procedure would yield similar results.

To conclude, even though EC resembles Pavlovian conditioning procedurally, several studies suggest that once a stimulus has acquired a valence as the result of being paired with a liked or disliked stimulus, this acquired valence might not (always) be easily changed by presenting the stimulus on its own. Importantly, the low extinction susceptibility of EC does not make conditioned changes in liking unchangeable. The literature reports two promising procedures for altering conditioning preferences: US-revaluation and CC. Surprisingly, both phenomena have received little attention in EC research. Moreover, the studies that have been conducted provide conflicting findings on the efficacy of these techniques for changing conditioned valence. Little is also known on the sustainability (e.g., over time and contexts) of the possible evaluative effects that both procedures might bring forth. Given the high potential use of techniques through which a person's existing (dis)likes can be altered, we believe it to be important that this gap in the literature is addressed.

The present dissertation: A deeper investigation of the effect and properties of counterconditioning in EC

The primary goal of the experimental studies conducted in the first line of research of the present dissertation was to help fill this gap in the literature by taking a closer look at the effect and properties of CC in EC. In this section we explain why we chose to work on CC and provide some more information on the main research aims.

Why counterconditioning?

The choice to focus on CC rather than US-revaluation was based on two grounds. First, despite the limited amount of research on the possible evaluative effects of a CC procedure, this technique is already frequently used in applied settings with the aim of changing people's existing evaluations. A common advertising strategy, for instance, is to pair a product or corporation that has acquired a negative connotation with positive messages/images in commercials or campaigns. Also in clinical practice, CC techniques are regularly employed (e.g., Korrelboom, van der Gaag, Hendriks, Huijbrechts, & Berretty, 2008). More experimental data to back up the assumption that human attitudes can be changed (in a durable manner) through CC therefore seemed timely. Second, CC might have a wider scope of application than US-revaluation. A disadvantage of the latter technique is that knowledge of the initial acquisition US is required for it to be applicable. In contrast, CC can also be used to target preferences with an unknown acquisition history. As mentioned in the General Introduction, fear conditioning research suggests that people can acquire fear for an originally neutral stimulus (e.g., an elevator) through its pairing with a threatening stimulus (e.g., a panic attack). In such a learning phase, the CS not only becomes a predictor for the US, but also acquires a negative connotation through EC. A common observation in clinical practice is that many phobics do not recall having had a traumatic experience in the presence of their phobic object (Mineka & Zinbarg, 2006). In such cases, it is unclear how US-revaluation could be applied to reduce the phobic object's acquired negative valence, whereas a CC approach is readily applicable.[9] Furthermore, even in cases where the original acquisition US is known, it might not always be easy to revaluate this stimulus/experience (e.g., a traumatic event). Pairing the CS with new positive experiences might then be more feasible.

Research aims

In a *first* step, we wanted to find a *robust EC paradigm* in which we could study CC. As mentioned earlier, several failures to observe or replicate EC effects have been reported in the literature (e.g., Field & Davey, 1999; Rozin et al., 1998). It is clear, however, that if one wants to investigate whether conditioned valence can be altered with a CC procedure, having a stable conditioning paradigm in which participants can learn new (dis)likes is a prerequisite. In a *second* step, we sought to examine whether we could *replicate* the finding that previously acquired conditioned valence can be altered by pairing the CS with a US that has a valence opposite to that of the original acquisition US. In a *third* step, we aimed at investigating some *properties* of CC that might prove important when one wants to use this technique in applied settings. More specifically, we wanted to examine its time- and context-sensitivity. In a *fourth*

[9] Note that this argument is less relevant for conditioned addictions as the US is generally known in these circumstances.

and final step, we wanted to test our hypothesis that return of fear can be reduced by presenting CC trials after extinction (see General Introduction).

Note that our third research question concerning the time- and context-sensitivity of evaluative CC effects can also provide insight into what is learned in such a procedure. In essence, a CC procedure can be conceptualized as a retroactive interference paradigm in which participants learn conflicting associations in two temporally separated phases. Studies on Pavlovian conditioning typically show that in such interference paradigms (like CC but also extinction), the second learning phase does not involve ‘unlearning’ but rather entails the learning of a new conditioned response that temporarily suppresses the former one. Evidence for this comes from studies on rapid reacquisition, spontaneous recovery, renewal, and reinstatement, phenomena that illustrate that the original conditioned response can quickly be restored with new learning trials or can suddenly re-emerge after the mere passage of time, a context change, or unsignaled presentations of the original US (for a review, see Bouton, 2002, 2004, all these phenomena have been observed after extinction and after CC). According to modern learning psychologists (e.g., Bouton, 2002), interference paradigms result in the creation of an ‘ambiguous’ CS. Temporal or physical context elements will then determine which association (i.e., the first- or second-learned) is retrieved and guides responding. Pavlovian conditioning studies further suggest that in such interference paradigms, second-learned information is more time- and context-dependent than first-learned information. According to Bouton (2002, 2004), our learning and memory system treats the first-learned information (e.g., CS-US or CS-US_1) as context-free, but the second-learned information (e.g., CS-noUS or CS-US_2) as a kind of time- and context-specific ‘exception to the rule’. An interesting question is whether the same applies for EC. Are second-learned conditioned evaluations able to overwrite previously acquired evaluations or do both evaluations (new and old) coexist in memory? The reappearance of participants’ initial evaluations after a period of time (i.e., spontaneous recovery) or after a context switch (i.e., renewal) would provide evidence for the latter alternative.

As previously mentioned, there is a lively debate in the EC literature on the processes and representations that may underlie EC effects (e.g., the automatic formation of associations/associations or propositional reasoning/ propositions). Remarkably, the different EC theories remain rather silent about the question of how second-learned evaluations might be represented in memory.[10] This question, has, however, received considerable attention in the domain of attitude research in social psychology.[11] Some authors support an overwriting view and believe that when an attitude changes, the old attitude disappears and is replaced with the new one. Anderson (1971, 1981), for instance, suggests that a person's new attitude is a weighted average of the old attitude and the new evaluative information/experiences to which the person is exposed. Upon change, the old attitude ceases to exist because its value has been incorporated into the new attitude position. A similar view is held by Fazio (2007), according to whom the new and old attitude are integrated into a unitary, univalent 'summary evaluation'. Other authors, on the other hand, argue that when attitude change occurs, the old evaluation does not disappear but rather continues to exist and can still have an impact on current evaluative responding. This view can, for instance, be recognized in the work of Gawronski and Bodenhausen (2006, 2007). In their APE (Associative-Propositional Evaluation) model, Gawronski and Bodenhausen argue that for each attitude object there is a network of evaluative associations in memory rather than one unitary summary evaluation. New evaluative experiences can create new associations in this network, but the old ones continue to exist as well. Other expressions of this view can be found in the Dual Attitude Model of Wilson, Lindsey, and Schooler (2000) and the Meta-Cognitive Model (MCM) and Past Attitudes Are Still There (PAST) model of Petty and colleagues (Petty, Briñol, & DeMarree, 2007; Petty, Tormala, Briñol, & Jarvis, 2006). Proponents of this vision all assume that rather than overwriting the old evaluation, new evaluative experiences result in a more complex, multifaceted evaluative representation. Different opinions exist, however, on when the old and/or newly acquired evaluation will influence responding. Some authors (e.g., Wilson et al., 2000;

[10] Probably because EC research has mainly focused on how evaluations can be *acquired* through EC. The question of how existing evaluations can be *changed* has received little attention (see also Walther & Langer, 2008).
[11] Note that the scope of attitude (change) models is much broader than attitude formation and change through EC.

Petty et al., 2006), for instance, assume that the initially acquired evaluation will influence more ‘automatic’ evaluative responding, that is, evaluative responding in situations with high time pressure or in which a person’s cognitive resources are limited. The newly acquired evaluation, on the other hand, will guide evaluative responding in more deliberative situations. Others claim that context cues might determine how an object is evaluated (e.g., Gawronski & Bodenhausen, 2006, 2007; Petty et al., 2006; Rydell & Gawronski, 2009). Evaluative responding in a certain context is then postulated to be congruent with the learning experiences that took place in that context. The latter approach comes close to Bouton’s (2002, 2004) earlier discussed theorizing on what is learned in Pavlovian retroactive interference paradigms (also see Rydell & Gawronski, 2009).

By investigating the time- and context-sensitivity of CC in EC, we hope to contribute to this debate and to help get a better understanding of what is learned in an evaluative retroactive CC interference paradigm. Note that, as mentioned, the answer to this question holds important implications for the use of CC techniques in applied settings. For instance, against the background of our hypothesis that CC might help to reduce (return of) fear, the finding that counterconditioned preferences would be time- and context-specific, would suggest that the possible beneficial therapeutic effects of a CC treatment might only be temporary. Participants’ original negative evaluation of the fear object might reappear over time or upon leaving the therapeutic context.

References

Anderson, N. H. (1971). Integration theory and attitude change. *Psychological Review, 78*(3), 171-206.

Anderson, N. H. (1981). *Foundations of information integration theory.* New York: Academic Press.

Baccus, J. R., Baldwin, M. W., & Packer, D. J. (2004). Increasing implicit self-esteem through classical conditioning. *Psychological Science, 15*, 498-502.

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation, 27*(2), 200-241.

Baeyens, F., Crombez, G., Hendrickx, H., & Eelen, P. (1995). Parameters of human evaluative flavor-flavor conditioning. *Learning and Motivation, 26*(2), 141-160.

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact - Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10*(4), 179-199.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning - Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*(2), 133-142.

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning - A case for unaware affective-evaluative learning. *Cognition & Emotion, 4*(1), 3-18.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective evaluative value - Conservative but not unchangeable. *Behaviour Research and Therapy, 27*(3), 279-287.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning - Acquired valence is sensitive to US-revaluation. *Learning and Motivation, 23*(2), 200-224.

Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither extended sequential nor simultaneous feature positive training result in modulation of evaluative flavor-flavor conditioning in humans. *Appetite, 31*(2), 185-204.

Baeyens, F., Vanhouche, W., Crombez, G., & Eelen, P. (1998). Human evaluative flavor-flavor conditioning is not sensitive to post-acquisition US-inflation. *Psychologica Belgica, 38*(2), 83-108.

Baeyens, F., Vansteenwegen, D., De Houwer, J., & Crombez, G. (1996). Observational conditioning of food valence in humans. *Appetite, 27*(3), 235-250.

Beckers, T., de Vicq, P., & Baeyens, F. (2009). Evaluative conditioning is insensitive to blocking. *Psychologica Belgica, 49*, 41-57.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 1-62). New York: Academic Press.

Blechert, J., Michael, T., Vriends, N., Margraf, J., & Wilhelm, F. H. (2007). Fear conditioning in posttraumatic stress disorder: Evidence for delayed extinction of autonomic, experiential, and behavioral responses. *Behaviour Research and Therapy, 45*, 2019-2033.

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation, 39*(1), 58-70.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*(10), 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*(5), 485-494.

Bouton, M. E., & Peck, C. A. (1992). Spontaneous recovery in cross-motivational transfer (counterconditioning). *Animal Learning & Behavior, 20*(4), 313-321.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes II: Arm flexion and extension have differential-effects on attitudes. *Journal of Personality and Social Psychology, 65*(1), 5-17.

Crano, W. D., & Prislin, R. (2006). Attitudes and persuasion. *Annual Review of Psychology, 57*, 345-374.

Craske, M. G., Hermans, D., & Vansteenwegen, D. (2006). *Fear and learning: Basic science to clinical application.* Washington, DC: American Psychological Association.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33*(2), 103-111.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology, 10*, 230-241.

De Houwer, J. (2009). How do people evaluate objects? A brief review. *Social and Personality Psychology Compass, 3*, 36-48.

De Houwer, J. (in press). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds.), *Applications of learning and conditioning*. Oxford, UK: Oxford University Press.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*(2), 161-174.

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology-Animal Behavior Processes, 26*(2), 237-242.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853-869.

Delamater, A. R., & LoLordo, V. M (1991). Event revaluation procedures and associative structures in Pavlovian conditioning. In L. Dachoswki & C. Flaherty (Eds.), *Current topics in animal learning: Brain, emotion, and cognition*. Hillsdale, N.J.: Erlbaum.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*(2), 245-268.

Dickinson, A., & Brown, K. J. (2007). Flavor-evaluative conditioning is unaffected by contingency knowledge during training with color-flavor compounds. *Learning & Behavior, 35*(1), 36-42.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*(2), 345-355.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*(5), 549-554.

Dwyer, D. M., Jarratt, F., & Dick, K. (2007). Evaluative conditioning with foods as CSs and body shapes as USs: No evidence for sex differences, extinction, or overshadowing. *Cognition & Emotion, 21*(2), 281-299.

Eifert, G. H., Craill, L., Carey, E., & O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour Research and Therapy, 26*(4), 321-330.

Fazio, R. H. (2007). Attitudes as object-evaluation associations of varying strength. *Social Cognition, 25*(5), 603-637.

Fazio, R. H., Eiser, J. R., & Shook, N. J. (2004). Attitude formation through exploration: Valence asymmetries. *Journal of Personality and Social Psychology, 87*(3), 293-311.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229-238.

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 161-202). New York: Academic Press.

Field, A. P. (2006). I don't like it because it eats sprouts: Conditioning preferences in children. *Behaviour Research and Therapy, 44*(3), 439-455.

Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology-Animal Behavior Processes, 25*(2), 211-224.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692-731.

Gawronski, B., & Bodenhausen, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition, 25*, 687-717.

Gibson, B. (2008). Can Evaluative conditioning change attitudes toward mature brands? New evidence from the implicit association test. *Journal of Consumer Research, 35*(1), 178-188.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Hamm, A. O., & Vaitl, D. (1996). Affective learning: Awareness and aversion. *Psychophysiology, 33*(6), 698-710.

Hammerl, M., Bloch, M., & Silverthorne, C. P. (1997). Effects of US-alone presentations on human evaluative conditioning. *Learning and Motivation, 28*(4), 491-509.

Hardwick, S. A., & Lipp, O. V. (2000). Modulation of affective learning: An occasion for evaluative conditioning? *Learning and Motivation, 31*(3), 251-271.

Hermans, D. (1998). Inleiding: Evaluatieve conditionering. *Gedragstherapie, 31,* 3-6.

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S. P. Shohov (Ed.), *Advances in Psychology Research* (pp. 17-41). Huntington, NY: Nova Science.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (in press). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*.

Holland, P. C. (1983). Occasion-setting in Pavlovian feature-positive discriminations. In M. L. Commons, R. J. Herrnstein & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Discrimination processes* (pp. 183-206). Cambridge, MA: Ballinger.

Holland, P. C. (1991). Acquisition and transfer of occasion setting in operant feature positive and feature negative discriminations. *Learning and Motivation, 22*(4), 366-387.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the implicit association test. *Journal of Personality and Social Psychology, 81*(5), 774-788.

Kerpelman, J. P., & Himmelfarb, S. Z. (1971). Partial reinforcement effects in attitude acquisition and counterconditioning. *Journal of Personality and Social Psychology, 19*(3), 301-305.

Korrelboom, C. W., van der Gaag, M., Hendriks, V. M., Huijbrechts, I. P. A. M., & Berretty, E. W. (2008). Treating obsessions with Competitive Memory Training: A pilot study. *The Behavior Therapist, 31*, 29-36.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*(3), 377-395.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures - Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261-273.

Levey, A. B., & Martin, I. (1975). Classical-conditioning of human evaluative responses. *Behaviour Research and Therapy, 13*(4), 221-226.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140*, 1-55.

Lipp, O. V., Neumann, D. L., & Mason, V. (2001). Stimulus competition in affective and relational learning. *Learning and Motivation, 32*(3), 306-331.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative Learning in human pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology-Animal Behavior Processes, 32*(4), 431-440.

Martin, I., & Levey, A. B. (1978). Evaluative conditioning. *Advances in Behaviour Research and Therapy*, *1*, 57-101.

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders - It's not what you thought it was. *American Psychologist, 61*(1), 10-26.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*(4), 421-433.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Chicago: University of Illinois Press.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation, 21*(1), 1-31.

Petty, R. E., Briñol, P., & DeMarree, K. G. (2007). The Meta-Cognitive Model (MCM) of attitudes: Implications for Attitude measurement, change, and strength. *Social Cognition, 25*, 657-686.

Petty, R. E., Tormala, Z. L., Briñol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology, 90*(1), 21-41.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology-Learning Memory and Cognition, 33*(1), 130-144.

Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation, 32*(1), 84-99.

Razran, G. (1954). The conditioned evocation of attitudes (cognitive conditioning?). *Journal of Experimental Psychology, 48*, 278-282.

Rescorla, R. A. (1985). Conditioned inhibition and facilitation. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 229-326). Hillsdale, NJ: Erlbaum.

Rescorla, R. A. (1991). Transfer of inhibition and facilitation mediated by the original target stimulus. *Animal Learning & Behavior, 19*(1), 65-70.

Rozin, P. (1982). Human food selection: The interaction of biology, culture and individual experience (pp. 225-254). In L. M. Barker (Ed.), *The psychobiology of human food selection*. Westport, CT, AVI Publishing.

Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes - A twin study. *Appetite, 8*(2), 125-134.

Rozin, P., Wrzesniewski, A., & Byrnes, D. (1998). The elusiveness of evaluative conditioning. *Learning and Motivation, 29*(4), 397-415.

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition & Emotion, 23*(6), 1118-1152.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*(6), 995-1008.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology, 37*, 867-878.

Sherer, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion, 7*, 1-42.

Staats, A. W., & Staats, C. K. (1957). Attitudes established by classical conditioning as measured by the semantic differential. *American Psychologist, 12*(4), 363.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of unconditioned stimulus valence and unconditioned stimulus identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology, 97*(3), 404-420.

Stevenson, R. J., Boakes, R. A., & Wilson, J. P. (2000). Counter-conditioning following human odor-taste and color-taste learning. *Learning and Motivation, 31*(2), 114-127.

Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (1998). Extinction in fear conditioning: Effects on startle modulation and evaluative self-reports. *Psychophysiology, 35*(6), 729-736.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology-Animal Behavior Processes, 32*(1), 71-79.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition & Emotion, 23*(5), 889-917.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology-Animal Behavior Processes, 32*(4), 454-459.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition & Emotion, 19*(2), 175-196.

Walther, E., & Langer, T. (2008). Attitude formation and change through association: An evaluative conditioning account. In: R. Prislin & W. B. Crano (Eds.), *Attitudes and attitude change* (pp. 87-109). Psychology Press, New York.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101-126.

Winkielman, P., Schwarz, N., Fazendeiro, T., & Reber, R. (2003). The hedonic marking of processing fluency: Implications for evaluative judgment. In J. Musch & K. C. Klauer (Eds.), *The Psychology of Evaluation: Affective Processes in Cognition and Emotion.* (pp. 189-217). Mahwah, NJ: Lawrence Erlbaum.

Woud, M. L., Becker, E. S., & Rinck, M. (2008). Implicit evaluation bias induced by approach and avoidance. *Cognition & Emotion, 22*(6), 1187-1197.

Zajonc, R. B. (2001). Mere exposure: A gateway to the subliminal. *Current Directions in Psychological Science, 10*(6), 224-228.

experiments 1-2

Finding a stable paradigm to study evaluative learning

In order to study whether conditioned valence can be altered with a counterconditioning procedure, a stable evaluative conditioning (EC) paradigm is needed. Experiments 1 and 2 were aimed at finding such a robust EC paradigm. In Experiment 1, a picture-picture procedure that was based on Walther, Gawronski, Blank, and Langer (2009) was tested. Experiment 2 was set up to test two gustatory EC procedures. Both studies consisted of three consecutive phases: 1) a pre-measurement of the valence of the CSs, 2) an evaluative conditioning phase in which two CSs were consistently paired with a positive US while two other CSs were always combined with a negative US, and 3) a post-measurement of CS valence. Self-report and affective priming results indicated strong EC effects in both gustatory paradigms of Experiment 2. In sharp contrast, the data of the picture-picture study revealed only a trend towards an EC effect.

Experiment 2 is published as: Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2009). A picture-flavor paradigm for studying complex conditioning processes in food preference learning. *Appetite, 53,* 303-308.

experiment 1

Introduction

Preference learning through evaluative conditioning (EC) has been demonstrated in a large number of studies using a variety of paradigms (e.g., a picture-picture paradigm, Pleyers, Corneille, Luminet, & Yzerbyt, 2007; a picture-odor procedure, Todrank, Byrnes, Wrzesniewski, & Rozin, 1995; a flavor-flavor paradigm, Baeyens, Eelen, Van den Bergh, & Crombez, 1990). Despite the fact that there is now general agreement in the field that EC is a genuine phenomenon (e.g., Hofmann, De Houwer, Perugini, Baeyens, & Crombez, in press), several failures to observe or replicate EC effects have been reported in the literature (e.g., Field & Davey, 1999; Rozin, Wrzesniewski, & Byrnes, 1998). Evaluative conditioning seems to be subject to many as yet unidentified boundary conditions (De Houwer, Baeyens, & Field, 2005). It is clear, however, that if one wants to investigate whether conditioned valence can be altered with a counterconditioning procedure, having a stable conditioning paradigm in which participants can learn new (dis)likes is a prerequisite. The main goal of the experiments presented in this chapter (Experiments 1-2) was to find such a robust EC paradigm.

In Experiment 1, a picture-picture procedure was put to the test in which both the CSs and USs were pictures of human faces. This procedure was based on the work of Walther, Gawronski, Blank, and Langer (2009), who recently obtained strong evaluative learning effects with this paradigm in several experiments. An important difference between Walther et al.'s procedure and other picture-picture procedures is that the USs are 'boosted' before conditioning. A prerequisite for EC effects to occur is that the applied USs are experienced as intended by the participants. In previous picture-picture studies, the USs were selected based on ratings made by individual participants (e.g., De Houwer, Baeyens, Vansteenwegen, & Eelen, 2000) or an independent group of raters (e.g., IAPS-pictures, Pleyers et al., 2007). Walther et al. (2009) applied the latter selection procedure, but further enhanced the valence of the pre-selected USs in a 'US formation phase'. In this formation phase, the US individuals were presented together with a number of either positive (e.g., 'is almost always in a good mood') or negative statements (e.g., 'is always very aggressive') about them, corresponding with their a-priori valence. The aim of this phase was to create a strong and stable US valence.

In the present experiment, we followed Walther et al. (2009) by including a US formation phase in which the valence of pre-selected USs (two positive; two negative) was boosted. In the subsequent EC phase, two neutral CSs were paired with a positive US and two other CSs were combined with a negative US. We expected the CSs to acquire the valence of the US with which they were presented contingently. Valence was assessed directly with ratings, but also indirectly with an Affective Priming Task (APT; Fazio, Sanbonmatsu, Powell, & Kardes, 1986). An important advantage of indirect measures like the APT is that they are less prone to demand bias when compared to ratings (De Houwer, 2006).

Method

Participants

Forty-two naive psychology students (32 women) participated for partial fulfillment of course requirements.

Materials

The stimulus material consisted of 18 black-and-white portrait photographs of male individuals. The majority of these pictures were selected on the basis of a rating study in which participants (N = 27) rated 38 photographs of male individuals for likeability on a scale ranging from -100 (*very unpleasant*) to +100 (*very pleasant*). The two pictures with the highest (M_{pos1} = 46.30, SD_{pos1} = 28.84; M_{pos2} = 57.78, SD_{pos2} = 25.01), respectively lowest (M_{neg1} = -70.37, SD_{neg1} = 27.52; M_{neg2} = -57.04, SD_{neg2} = 25.69) scores were selected as positive, respectively negative, USs. The nine most neutral pictures (*Ms* between 0.37 and 20.37, *SDs* between 30.82 and 42.50) were chosen as 'potential' CSs. Five additional potential CS-pictures were obtained from the internet. From this total set of 14 potential CSs, four CSs were selected on an individual basis (see

Procedure section). All photographs had a width of 344 pixels and a height of 427 pixels.

For the APT, the four CS-pictures served as primes, and 10 positive and 10 negative Dutch nouns (selected from Hermans & De Houwer, 1994) were used as targets.

An Affect 4.0 program (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010), run on Pentium IV computers, controlled stimulus presentation and response registration.

Procedure

Participants were run individually in 10 adjacent isolated cubicles. The experimental procedure was modeled after Walther et al. (2009). Participants were greeted by the experimenter and seated in front of a computer screen. The experiment consisted of four sequential phases, which were guided entirely by a computer program: a pre-acquisition valence measurement, a US formation phase, an EC phase and a post-acquisition valence measurement.

<u>*Pre-acquisition valence measurement*</u>. The experiment started with a baseline valence measurement during which participants were asked to rate the 14 potential CS-pictures on a 21-point scale ranging from -100 (*very unpleasant*) to +100 (*very pleasant*). Based on these ratings, the computer program selected the four most neutral pictures for each participant to serve as CSs. Subsequently, valence was assessed by means of an APT. The experimenter explained that on each trial a picture (called the ‘prime’) would precede a word (called the ‘target’). Participants were instructed to attend the word and evaluate it as quickly as possible by pressing the right, respectively left, mouse button for positive, respectively negative, words. The priming task consisted of two blocks of 40 experimental trials, preceded by 12 practice trials. Each trial started with a fixation cross (500 ms), followed by a 500 ms blank screen, followed by the prime (200 ms). Fifty ms after prime offset, the target was presented until participants responded or 2000 ms elapsed. The mean intertrial interval was 1000 ms (range 500-1500 ms). Within each block every prime was paired with five negative and five positive targets. Each target appeared equally often in each block. The order of the trials within a block was randomized with the

restriction that the same prime should not be presented on more than two consecutive trials and that two successive trials should not contain the same target.

US formation phase. In the US formation phase, participants were asked to imagine that they had just started a new job in a company, and hence were interested in getting to know their new colleagues. Participants were then presented with the photographs of the four US individuals and a number of either positive or negative statements about these persons. The two positive USs were each paired with three positive statements (e.g., 'is always friendly'), while the two negative USs were combined with three negative statements (e.g., 'often comes to work drunk').[12] The photographs were presented on the left side of the screen and the statements (one at a time) on the right side. Participants' task was to form an impression of these individuals based on the statements. Picture-statement pairs were randomly presented one-by-one for 7000 ms with an intertrial interval of 1000 ms.

Evaluative conditioning phase. During the EC phase, participants were asked to imagine that they were now familiar with some of their new colleagues, but not with others. Participants were then presented with pairs of neutral, unfamiliar individuals (CSs) and familiar individuals from the attitude formation phase (USs). CS-US pairs were constructed randomly for each participant. Two of the four CSs were paired with a positive US while the other two were combined with a negative US. Each CS-US pair was presented 12 times during conditioning. The CS individuals were always presented on the left side of the screen while the US individuals appeared on the right side. On each trial, the CS preceded the US by 750 ms and overlapped with it for another 2500 ms. The US then remained alone on the screen for another 750 ms. The intertrial interval was 2000 ms. Trial order was randomized with the restriction that no CS-US pair could appear more than twice in a row. Participants' task was again to form an impression of the individuals presented on the screen.

[12] Other positive statements were: 'always listens very carefully', 'likes to help new colleagues to incorporate', 'is almost always in a good mood', 'is always there for colleagues when they need help', 'sees the positive in each situation'. Other negative statements were: 'is always very aggressive', 'often bothers other people', 'is almost always in a bad mood', 'becomes angry if people don't agree with him', 'often gossips about other people'.

Post-acquisition valence measurement. After the EC phase participants were asked to provide pleasantness ratings for all CSs and USs and completed the APT for a second time. Task order (priming-rating or rating-priming) was counterbalanced across participants. Finally, contingency recall was assessed. Each of the four CS individuals was presented with the four US pictures and participants were asked to indicate with which one it had been paired and how certain they were of this (*very uncertain*, *reasonably uncertain*, *reasonably certain*, *very certain*).

Results

Preliminary analyses

For each participant, we calculated the mean rating for CSs that were paired with a positive US, for CSs that were paired with a negative US, and for the positive and negative USs. These means were calculated separately for the pre-acquisition and post-acquisition ratings.

The priming data were corrected for outliers by excluding reaction times below 200 ms or above 1500 ms (1.06%). Trials with no (0.33%) or incorrect responses (8.54%) were also discarded. To create an evaluation score, the mean latency for positive target words was subtracted from the mean latency for negative words for each CS category (CSs paired with positive vs. negative US) (Gawronski, Walther, & Blank, 2005). Higher scores on this variable indicate more positive evaluations. Note that these scores should not be interpreted in an absolute manner (e.g., a value of zero reflecting a neutral evaluation), because response latencies for positive target words may generally differ from response latencies for negative target words.

US ratings

A prerequisite for EC effects to occur is that participants experience the used USs as intended. The analysis of the US ratings revealed that this precondition was met. As expected, the intended positive USs were rated significantly more positive (M = 63.81, SD = 25.82) than the intended negative USs (M = -61.19, SD = 27.34), $t(41) = 16.36$, $p < .0001$.

CS ratings

We expected that the CSs that were paired with a positive US during the EC phase would acquire a positive connotation while the CSs that were paired with a negative US would become negative. To test this hypothesis, a 2 (Moment: pre- vs. post-conditioning) x 2 (US-type: intended positive vs. negative) repeated measures ANOVA was conducted on the CS ratings.

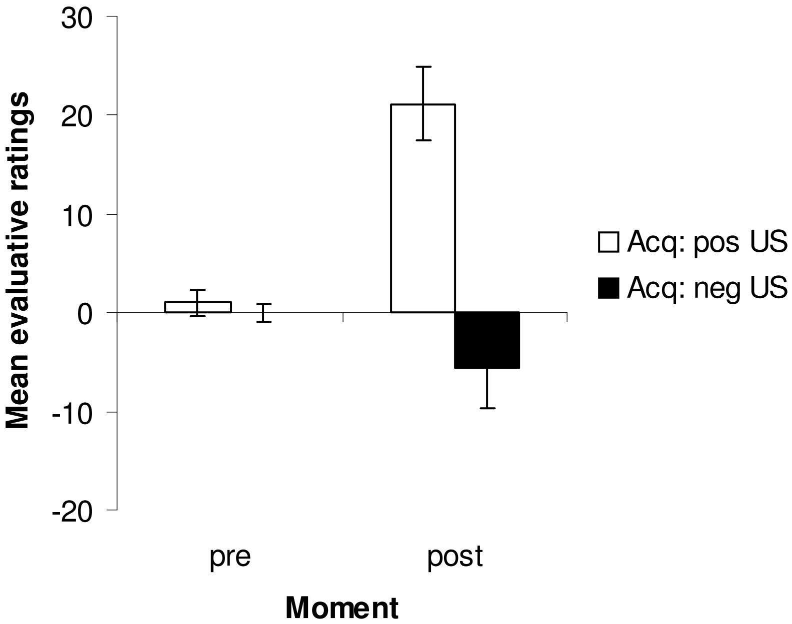


Figure 1. Mean evaluative ratings of the CSs that were paired with a positive (Acq: pos US) or negative US (Acq: neg US) during conditioning as a function of moment (Pre: before conditioning, Post: after conditioning). Error bars represent standard errors.

The crucial Moment x US-type interaction was significant, $F(1, 41) = 16.70$, $p < .001$.[13] Planned comparisons revealed that, as expected, CSs that were repeatedly paired with a positive US were evaluated more positively after ($M = 21.19$, $SD = 24.11$) as compared to before ($M = 0.95$, $SD = 8.35$) conditioning, $F(1, 41) = 219.48$, $p < .0001$. For the CSs that were paired with a negative US during conditioning, a trend was observed for a more negative evaluation after ($M = -5.60$, $SD = 26.25$) as compared to before ($M = 0.00$, $SD =$

[13] The analysis also revealed a main effect of moment, $F(1, 41) = 17.39$, $p < .001$, with more positive CS ratings after ($M = 7.80$, $SD = 12.83$) as compared to before conditioning ($M = 0.48$, $SD = 5.53$). Furthermore, a main effect of US-type was obtained, $F(1, 41) = 14.34$, $p < .001$. This reflects the fact that over moment, CSs that were paired with a liked US were rated more positively ($M = 11.07$, $SD = 14.69$) than CSs that were paired with a negative US ($M = -2.80$, $SD = 14.06$).

5.74) conditioning, but this difference was not significant, $F(1, 41) = 2.01$, *ns*. In the post-acquisition test, a significant difference was obtained between the CSs that had been paired with a positive US and the CSs that were combined with a negative US, $F(1, 41) = 16.01$, $p < .001$. This difference was not present in the pre-acquisition test, $F < 1$. The previous results make clear that the significant difference obtained in the post-acquisition test was mainly driven by the EC effect for the CSs that were paired with a positive US. The mean CS ratings before and after conditioning are depicted in Figure 1.

Priming data

For the priming data we also expected to observe an increase, respectively decrease, in positivity from pre- to post-test for the CSs that were paired with a positive, respectively negative US. A 2 (Moment: pre- vs. post-conditioning) x 2 (Block: 1 vs. 2) x 2 (US-type: intended positive vs. negative) repeated measures ANOVA was conducted on the priming scores. The crucial Moment x US-type interaction was not significant, $F < 1$, as were all other main and interaction effects. Figure 2 depicts the priming scores as a function of US-type and moment.

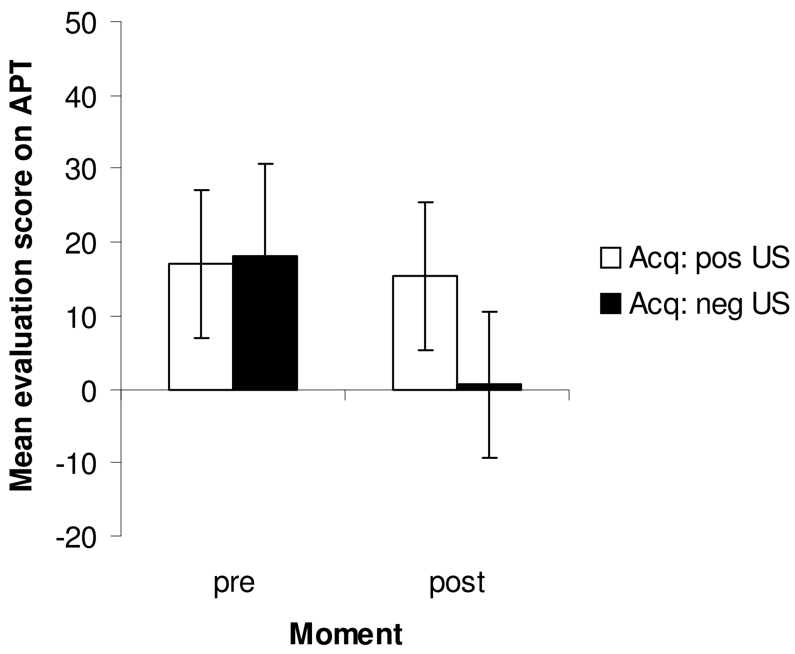


Figure 2. Mean evaluation score on the APT for CSs that received acquisition trials with a positive (Acq: pos US) or negative US (Acq: neg US) as a function of moment (Pre: before conditioning, Post: after conditioning). Error bars represent standard errors.

In contrast with our predictions, Figure 2 shows a similar evaluation score before (*M* = 17.03, *SD* = 65.58) and after (*M* = 15.39, *SD* = 65.10) conditioning for the CSs that were paired with a positive US, $F < 1$. For the CSs that were paired with a negative US, Figure 2 illustrates a decrease in positivity from pre-conditioning (*M* = 18.12, *SD* = 81.49) to post-conditioning (*M* = 0.64, *SD* = 64.20). This decrease failed to approach significance, however, $F(1, 41) = 1.36$, *ns*. In line with expectations, higher post-test scores were observed for the CSs that were paired with a positive US as compared to the CSs that were combined with a negative US. This difference was, however, also not significant, $F(1, 41) = 1.26$, *ns*.

In sum, no strong EC effects were observed in the priming data. The priming scores showed a trend towards a decrease in positivity for the CSs that were paired with a negative US, but no effect for the CSs that were combined with a positive US. This contrasts with the results of the ratings which revealed a significant effect for the CSs that were paired with a positive US but not for the CSs that were followed by a negative US.

Contingency recall

A participant was considered to remember a certain CS-US association when he/she correctly identified the US that had been paired with a specific CS, and was at least ‘reasonably certain’ of his/her choice. The mean number of recalled pairings was 2.83 (*SD* = 1.27, range 0-4) out of a maximum of four. Hence, participants remembered a considerable number of experimental contingencies at the end of the experiment.

Discussion

The aim of this experiment was to examine whether EC effects could be obtained with a picture-picture procedure that was modeled after Walther et al. (2009). A significant EC effect was observed in the ratings for the CSs that were paired with a positive US, but not for the CSs that were combined with a negative US. The analysis of the APT data yielded no significant results at all. The priming scores showed a trend towards a decrease in positivity for the CSs that were paired with a negative US, but no effect for the CSs that were combined with a positive US. Hence, a somewhat conflicting pattern of results was obtained for the rating and priming data.

It is not clear why we failed to replicate the EC effects obtained by Walther et al. (2009). A possible explanation might be found in two - at first sight seemingly small - alterations we made to their procedure. The first alteration concerns the number of acquisition trials. Whereas in the Walther et al. (2009) study each CS-US pair was presented five times during the EC phase, we increased the number of CS-US pairings to 12 based on the assumption that more acquisition trials would lead to stronger EC effects. Also previous studies conducted with the picture-picture paradigm in our and other labs commonly used 10 or more acquisition trials (e.g., Baeyens, Crombez, Van den Bergh, & Eelen, 1988; Baeyens, Eelen, & Van den Bergh, 1990; Baeyens, Eelen, Van den Bergh, & Crombez, 1989a, 1989b; Field & Moore, 2005; Lascelles & Davey, 2006; Olson & Fazio, 2006). Upon further investigation of the available evidence concerning the impact of the number of acquisition trials on EC effects, it became clear that, despite its obvious relevance to EC, only a limited number of studies have looked at this variable (see De Houwer, Thomas, & Baeyens, 2001). In addition, the few studies that did manipulate the number of conditioning trials found conflicting results. Some studies observed an increase in the magnitude of EC effects when the number of pairings increased (Baeyens, Eelen, Crombez, & Van den Bergh, 1992; Bar-Anan, De Houwer, & Nosek, 2009; Sachs, 1975; Staats & Staats, 1959), but others found no effect (Martin & Levey, 1987; Stuart, Shimp, & Engle, 1987). One study even suggested that EC effects might decrease in magnitude after the number of pairings surpasses 10 (Baeyens et al., 1992). Given the limited and mixed findings regarding the impact of the number of acquisition trials on EC, it is difficult to judge whether this factor might have contributed to the weak EC effects observed in our study.

The second alteration relates to the presentation of the CS and US. In Walther et al.'s (2009) study, the CS and US were presented simultaneously for four seconds. In the present experiment, a delayed conditioning procedure was used in which the CS preceded the US by 750 ms and overlapped with it for another 2500 ms. Hence, both the temporal sequence of the CS and US and their duration differed in our study. We opted for a delayed conditioning procedure to ensure that participants would pay attention to both the CS and the US. A wide range of presentation parameters has been used in the EC literature (e.g., delayed conditioning, Walther, 2002, Experiment 4; trace conditioning, Field & Moore, 2005; backward conditioning, Stuart et al., 1987; CS and US durations ranging between 1000 and 7000 ms, Field, 2000; Lascelles, Field, & Davey, 2003). A recent meta-analysis by Hofmann et al. (in press) demonstrated that the strength of EC effects was not moderated by the temporal sequence of the CS-US pairings (forward, backward or simultaneous) nor by the duration of the CS and US. These findings suggest that it is unlikely that the difference in results between our study and that of Walther et al. (2009) can be attributed to these procedural factors.

Based on the current evidence in the EC literature, it is not clear whether (and why and how) the small alterations we made to the procedure of Walther et al. (2009) are responsible for our failure to replicate their findings. A host of other factors that are difficult to compare might have played a role (e.g., attentional factors, Pleyers, Corneille, Yzerbyt, & Luminet, 2009; contingency awareness, Pleyers et al., 2007). Furthermore, we cannot explain why the rating and priming data showed somewhat conflicting results.

This experiment is not the first failure to replicate EC effects (e.g., Field & Davey, 1999; Rozin et al., 1998). As mentioned in the introduction, EC seems to be subject to many as yet unidentified boundary conditions. A lot more research is needed to explore and delineate these conditions. The present dissertation was not aimed at investigating the boundary conditions of EC. The goal of the experiments reported in this chapter was to find a stable EC paradigm that would allow us to study evaluative counterconditioning in a next step. The picture-picture procedure that was applied in the present experiment yielded only weak conditioning effects and therefore seems inappropriate for this purpose. Because successful EC effects were obtained in the other two paradigms we tested[14] (see Experiment 2), we did not further explore whether we could acquire stronger EC effects with the picture-picture paradigm by making some alterations to it (e.g., reducing the number of acquisition trials).

[14] Experiment 1 and 2 were conducted at the same time.

References

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact always in contact - Evaluative conditioning is resistant to extinction. *Advances in Behaviour Research and Therapy, 10*(4), 179-199.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning - Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*(2), 133-142.

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning - A case for unaware affective-evaluative learning. *Cognition & Emotion, 4*(1), 3-18.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989a). Acquired affective evaluative value - Conservative but not unchangeable. *Behaviour Research and Therapy, 27*(3), 279-287.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989b). The influence of CS-UCS perceptual similarity/dissimilarity on human evaluative learning and signal learning. *Learning and Motivation, 20*(3), 322-333.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation, 21,* 434-455.

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2009). The role of contingency memory, intentional processes and number of pairings in evaluative conditioning. Unpublished Manuscript.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA: Sage Publishers.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*(2), 161-174.

De Houwer, J., Baeyens, F., Vansteenwegen, D., & Eelen, P. (2000). Evaluative conditioning in the picture-picture paradigm with random assignment of conditioned stimuli to unconditioned stimuli. *Journal of Experimental Psychology - Animal Behavior Processes, 26*(2), 237-242.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853-869.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229-238.

Field, A. P. (2000). I like it, but I'm not sure why: Can evaluative conditioning occur without conscious awareness? *Consciousness and Cognition, 9*(1), 13-36.

Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology - Animal Behavior Processes, 25*(2), 211-224.

Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion, 19*(2), 217-243.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology, 41*(6), 618-626.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34*, 115-139.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (in press). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*.

Lascelles, K. R. R., & Davey, G. C. L. (2006). Successful differential evaluative conditioning using simultaneous and trace conditioning procedures in the picture-picture paradigm. *Quarterly Journal of Experimental Psychology, 59*(3), 482-492.

Lascelles, K. R. R., Field, A. P., & Davey, G. C. L. (2003). Using foods as CSs and body shapes as UCSs: A putative role for associative learning in the development of eating disorders. *Behavior Therapy, 34*(2), 213-235.

Martin, I., & Levey, A. (1987). Relation between verbalizable knowledge and conditioned responding. *Bulletin of the British Psychological Society, 40*, A2-A3.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*(4), 421-433.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology - Learning Memory and Cognition, 33*(1), 130-144.

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology - Animal Behavior Processes, 35*(2), 279-285.

Rozin, P., Wrzesniewski, A., & Byrnes, D. (1998). The elusiveness of evaluative conditioning. *Learning and Motivation, 29*(4), 397-415.

Sachs, D. H. (1975). Conditioning of affect to a neutral stimulus - Number of trials. *Perceptual and Motor Skills, 40*(3), 895-901.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45.

Staats, A. W., & Staats, C. K. (1959). Effect of number of trials on the language conditioning of meaning. *Journal of General Psychology, 61*(2), 211-223.

Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical-conditioning of consumer attitudes - 4 experiments in an advertising context. *Journal of Consumer Research, 14*(3), 334-349.

Todrank, J., Byrnes, D., Wrzesniewski, A., & Rozin, P. (1995). Odors can change preferences for people in photographs - A cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation, 26*(2), 116-140.

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82*(6), 919-934.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition & Emotion, 23*(5), 889-917.

experiment 2

Introduction[15]

Human food choice has been described as a complex process, influenced by numerous variables (e.g., Eertmans, Baeyens, & Van den Bergh, 2001). Apart from factors such as cost and availability, acceptance or rejection of a food may be based on expectations about the short- or long-term beneficial or harmful consequences of consuming it. An even more important role, however, is attributed to the affective reaction to the sensory characteristics (e.g., flavor, texture, color) of the food stimulus, in other words to liking or disliking it (Eertmans et al., 2001). In general, humans tend to eat foods they like and turn down foods they do not like. Hence, to learn more about human food choice and intake, understanding how food (dis)likes are formed and changed is crucial.

Empirical evidence suggests that in humans, most of the evaluative reactions towards foods are learned rather than innate (Rozin & Millman, 1987). Several studies, moreover, suggest that an important route through which food (dis)likes can be acquired is evaluative conditioning (EC). EC refers to changes in the evaluation of a stimulus that result from pairing it with other positive or negative stimuli (De Houwer, Thomas, & Baeyens, 2001). In a prototypical EC experiment, a neutral stimulus (the Conditioned Stimulus or CS) is presented repeatedly with a subjectively liked or disliked stimulus (the Unconditioned Stimulus or US). What is typically observed is that after these repeated CS-US pairings, the CS valence shifts in the direction of the US valence. Hence, a new preference is learned. Hermans, Baeyens, Lamote, Spruyt and Eelen (2005), for example, induced food (dis)likes by applying an EC procedure in which pictures of yoghurt brands served as CSs and odors as USs. Yoghurt brands that were paired with a positive odor acquired a positive connotation, while brands paired with a negative odor became negative. An example closer to daily life would be a child that learns to like orange flavor as a consequence of its repeated co-occurrence with the agreeable sweetness of sugar in the average orange soft drink.

[15] Because this study was submitted to *Appetite*, strong emphasis is put on the relevance of our findings for the food literature.

Several authors have pointed out that the EC paradigm provides an interesting framework for investigating the formation and change of (dis)likes under experimentally controlled conditions (De Houwer et al., 2001). An obstacle, however, is that EC effects appear to be not very robust. The literature reports several failures to find EC effects (e.g., De Houwer et al., 2001). Identifying possible boundary conditions and developing robust paradigms for studying EC therefore form two important pathways in current research. In the context of the latter pathway, an interesting paradigm for studying food (dis)likes was developed by Verhulst, Hermans, Baeyens, Spruyt, and Eelen (2006). These authors applied an EC procedure in which real but unknown food items - small cookies differing in color and shape - were paired with positive or negative flavors. They found that cookies that were repeatedly paired with a positive flavor acquired a positive valence whereas cookies that were combined with a negative flavor acquired a negative valence. This EC effect was not only evidenced by ratings, but also by data from an indirect measure, namely the Affective Priming Task (APT). This is an important strength of the Verhulst et al. (2006) study as in the majority of published studies on food likes, food preferences were only assessed by direct, verbal methods. A significant limitation of relying solely on verbal report is that the data can be flawed by social desirability or self-presentation biases. Indirect measures of attitudes/liking are assumed to be less influenced by such concerns and therefore form an important alternative or complement to direct measures in areas like food preference, where social desirability or normative pressures can be at play. In several unpublished studies in our lab (for an overview, see Verhulst, 2007) the results of Verhulst et al. (2006) were replicated. Hence, their cookie-paradigm constitutes a robust and promising tool to learn more about food (dis)likes.

Traditionally, studies on EC of food (dis)likes have focused on simple acquisition processes (e.g., Hermans et al., 2005; Verhulst et al., 2006). Recently, however, interest has grown in the role of more complex conditioning processes (e.g., extinction, blocking, latent inhibition) in preference learning. For instance, an interesting question with respect to the child that learned to like orange flavor through its pairing with sugar, would be whether this acquired preference can be abolished by presenting the orange flavor repeatedly on its own (i.e., without sugar). This parallels the question of whether conditioned preferences can be altered through an extinction procedure. Another question would be whether a child who earlier experienced orange flavor without sugar would just as easily acquire a preference for this flavor when it is paired with sugar. This corresponds to the question of whether conditioned preference learning is susceptible to latent inhibition.

Although the procedure described by Verhulst and colleagues (2006) seems to be a promising paradigm to investigate these more complex conditioning processes, it has a serious limitation. As indicated earlier, in this procedure, the color-shape compound of the cookies served as CS and the positive or negative flavor as US. Color, shape and flavor were hereby all combined in one cookie, implying that the CS was part of a complex stimulus configuration rather than a separate autonomous stimulus (i.e., CS and US were part of the same stimulus). This makes the cookie-paradigm of Verhulst et al. (2006) less suitable for research questions that require separate CS or US presentations (e.g., the mentioned questions on extinction and cue competition) and hence considerably limits its applicability.

To deal with this limitation, an adapted version was developed for the present experiment in which the CS and US no longer form a compound. Here, the USs are still positively and negatively flavored food items, but the CSs are pictures of cookies instead of real color-shape compounds. Hence, CS and US are physically separated. This separation provides a solution for the problem described above, but it is unclear how it will impact the conditioning effects. One consequence of the procedural change could be that the procedure becomes more artificial as a mental distance arises between the CS-picture and the US cookie, i.e., the CS becomes a more distal cue. Research indeed suggests that developing an associative CS-US link is easier when they show some sort of ‘belongingness’ or ‘fit’ (Hamm, Vaitl, & Lang, 1989). In addition, typically more proximal cue procedures are used in EC research. Another consequence could be that CS-US contingency awareness increases. At least some authors argue that contingency awareness is necessary for EC effects to occur (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007 but see Walther & Nagengast, 2006 for alternative findings). In contrast to the Verhulst et al. (2006) procedure, where participants are no longer able to see the color-shape CS when they eat the cookie and experience the US, the new procedure entails that the CS characteristics are still present when the US is presented. Both factors might influence the obtained evaluative learning effects. Therefore, in the current study, we compared our adapted cookie-paradigm to the procedure described by Verhulst et al. (2006).

Similar to Verhulst et al. (2006), valence was assessed both directly (ratings) and indirectly (APT). A new element of the present study was that the APT was applied not only *after*, but also *before* conditioning took place (i.e., as an indirect pre/post measurement). Previous studies (e.g., Lamote, Hermans, Baeyens, & Eelen, 2004; Verhulst et al., 2006) found the APT capable of measuring long existing and recently acquired food attitudes. The present study explored the use of the APT as indirect measure of *changes* in food attitudes over time.

Method

Participants

Ninety-six naive psychology students (75 women) participated for partial fulfillment of course requirements (72 students) or a monetary reward of € 5 (24 students). Half of them were assigned to the procedure described by Verhulst et al. (2006) (the '*compound* presentation group'), whereas the other half received the adapted procedure (the '*separate* presentation group').

Materials

For the compound presentation group, we designed four cookies differing in color and shape (see Verhulst et al., 2006). They had a diameter of approximately 1 cm and were a light yellow triangle, a red moon shape, a green star and a dark yellow pear shape. Pastry for the cookies was made out of 200 g of flour, 100 g of sugar ('intended good tasting' cookies only), one egg and 125 g of butter. The pastry was divided in four equal parts and by adding a tablespoon of hazelnut powder or honey 'intended good tasting cookies' (~ 8 kcal/cookie) were obtained. Addition of vegetable bouillon or two tablespoons of Tween20 (i.e., *polysorbate 20*) gave rise to 'intended bad tasting cookies' (~ 5 kcal/cookie). The pairing of flavors and color-shape compounds was counterbalanced over participants.

For the separate presentation group, two intended good tasting (hazelnut or honey) and bad tasting (bouillon or Tween20) cookies were designed as well. An important difference, however, was that these cookies no longer differed in color and shape. All cookies were yellow squares with a diameter of approximately 1 cm. Pictures of the four cookies of the compound presentation group served as CSs. These were centered on a grey background with a displayed diameter of 5-6 cm. The total picture was 512 by 384 pixels. Also here, CS-US pairings were fully counterbalanced across participants.

For the priming task, pictures of the cookies of the compound presentation group served as primes. Targets were ten positive and ten negative Dutch nouns selected from Hermans and De Houwer (1994).

An Affect 4.0 program (Hermans, Clarysse, Baeyens, & Spruyt, 2005) controlled the presentation of all stimuli, as well as the registration of all responses. The experiment was run on IBM compatible Pentium IV computers.

Procedure

Participants were run individually in ten adjacent isolated cubicles. The experiment consisted of a pre-acquisition valence measurement, an EC phase and a post-acquisition valence measurement.

The experiment started with a baseline valence measurement during which participants rated each cookie-picture by shifting a pointer on a 21-point scale ranging from -100 (*very unpleasant)* to +100 (*very pleasant*). Participants were encouraged to rely on their first, spontaneous reaction toward the pictures. Valence was also assessed with the APT. The experimenter explained that pairs of stimuli would be presented on the computer screen, the first of which would always be a picture (the prime), whereas the second would always be a word (the target). Participants were instructed to attend to the word and to evaluate it as quickly as possible by pressing the left mouse button for negative words and the right button for positive words. The priming task consisted of two blocks of 40 experimental trials, preceded by 12 practice trials. Each trial started with a fixation cross (500 ms), followed by a 500 ms blank screen, followed by the presentation of the prime for 200 ms. Fifty ms after prime offset, the target was presented until the participant responded or 2000 ms elapsed. The intertrial interval varied from 500 to 1500 ms with a mean of 1000 ms. In each block each prime was paired with five negative and five positive targets. Prime-target combinations were selected so that each target appeared equally often in each block. The order of the trials within a block was randomized with the restriction that the same prime should not be presented on more than two consecutive trials and that two successive trials should not contain the same target. Task order (priming-rating vs. rating-priming) was counterbalanced across participants.

The EC phase consisted of 24 trials. In the compound presentation group, participants were instructed to eat the 24 cookies one by one. Each cookie was presented six times. Presentation order was randomized for every participant, with the restriction that the same cookie could not be presented twice in a row. A dish filled with small pieces of bread was available for each participant and they were instructed to eat a piece between each trial to avoid aftertastes. Participants in the separate presentation group received four columns of all similar looking cookies. Each column contained cookies with the same flavor and above each column was a small picture of one of the CSs. The CS-pictures were presented one by one on the computer screen and participants were asked to eat a cookie from the column that had that same picture presented above it. All other presentation parameters were identical to those in the compound presentation group. For both experimental groups, the EC phase took approximately 12 minutes.[16]

After the EC phase participants were again asked to provide pleasantness ratings for the cookie-pictures and completed the priming task for a second time.

At the end of the experiment contingency knowledge was assessed by asking participants to indicate for each cookie-picture whether it had been paired with / followed by a hazelnut, honey, bouillon or Tween20 flavor and how certain

[16] Given this brief duration it is unlikely that differences in the caloric density or post-ingestive effects of the positive and negative cookies could have contributed to the outcome.

they were of this *(very uncertain, reasonably uncertain, reasonably certain, very certain)*. Additionally, participants rated the US flavors on a 21-point scale ranging from -100 *(very unpleasant)* to +100 *(very pleasant)*.

Results

Preliminary analyses indicated no effect of task order (rating-priming vs. priming-rating). This factor was therefore excluded from further analyses.

Two participants who failed to comply with task instructions were excluded from the priming analyses. The remaining data were corrected for outliers by excluding reaction times below 200 ms or above 1500 ms (.88 %). Trials on which no (.34%) or an incorrect response (6.91%) was given were also discarded.

Post-conditioning contingency knowledge

Participants were considered ‘contingency aware’ if they could correctly identify two out of four CS-US contingencies and were at least ‘reasonably certain’ of their choice. Six participants of the compound presentation group and nine of the separate presentation group were classified as ‘unaware’. A Chi-square test indicated no difference in the proportion of ‘aware’ participants between both experimental groups, $\chi^2(1) = .711$, *ns.* All participants were included in the subsequent analyses. Exclusion of ‘unaware’ participants did not affect the results.

US ratings

The mean US ratings were analyzed using a 2 x 2 ANOVA with presentation (compound vs. separate) as between-subjects variable and US-type (intended positive vs. negative) as within-subjects variable. A main effect of US-type was obtained, $F(1, 94) = 1057.40$, $p < .0001$, indicating that the USs had been rated as expected, with the intended positive USs rated significantly more positive ($M = 60.52$, $SD = 30.95$) than the intended negative USs ($M = -75.78$, $SD = 24.54$). The main effect of presentation and the Presentation x US-type interaction were not significant. Hence, if greater conditioning effects are obtained in one of the experimental groups, this cannot be due to a difference in US ratings.

Table 1

Means and standard errors (a) for the CS ratings and (b) response times to positive and negative targets before (Pre) and after (Post) the EC phase for both experimental groups

a	**Pos US Pre**		**Pos US Post**		**Neg US Pre**		**Neg US Post**	
	M	***SE***	***M***	***SE***	***M***	***SE***	***M***	***SE***
Compound	10.42	3.67	49.27	4.43	-0.73	3.94	-67.40	3.76
Separate	3.86	3.44	41.04	4.75	-5.52	6.61	-52.29	5.36

b	**Pos US Pre**				**Pos US Post**				**Neg US Pre**				**Neg US Post**			
	Pos T		**Neg T**		**Pos T**		**Neg T**		**Pos T**		**Neg T**		**Pos T**		**Neg T**	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Compound	640	13	664	13	619	15	656	14	647	14	644	12	653	14	625	13
Separate	635	13	630	13	591	15	620	14	630	14	636	12	623	14	602	13

Note. Pos/Neg US = Positive/Negative Unconditioned Stimulus, Pos/Neg T = Positive/Negative Target.

A closer inspection of the US ratings revealed that 91 out of 96 participants rated the positive USs as intended (i.e., as positive) and 95 out of 96 rated the negative USs as intended. Exclusion of participants who did not rate the USs as intended did not affect the results.

CS ratings

A 2 (Presentation: compound vs. separate) x 2 (Time: pre vs. post-conditioning) x 2 (US-type: intended positive vs. negative) ANOVA with repeated measures on US-type and time, was conducted on the CS ratings. The crucial Time x US-type interaction was significant, $F(1, 94) = 269.65$, $p < .0001$. In testing for simple effects, planned comparisons revealed that, as expected, CSs that were repeatedly paired with a positive flavor were afterwards evaluated more positively, $F(1, 95) = 113.24$, $p < .0001$, whereas CSs that were combined with a negative flavor acquired a negative valence, $F(1, 95) = 218.80$, $p < .0001$. At post-test, a significant difference was observed between CSs that had been paired with a positive ($M = 45.16$, $SD = 31.94$) versus a negative US ($M = -59.84$, $SD = 32.81$), $F(1, 95) = 387.92$, $p < .0001$. Unexpectedly, there was also a significant difference between CSs that would be paired with positive ($M = 7.14$, $SD = 24.74$) versus negative USs ($M = -3.13$, $SD = 26.17$) *before* EC took place, $F(1, 95) = 7.48$, $p < .05$. The significant Time x US-type interaction indicates, however, that this difference was far more pronounced *after* EC. Further planned contrasts revealed that the interaction between time and US-type was significant in both experimental groups, with $F(1, 47) = 174.79$, $p < .0001$ for the compound presentation group and $F(1, 47) = 101.51$, $p < .0001$ for the separate presentation group (see Table 1a). The three-way interaction between presentation, time and US-type was only marginally significant, $F(1, 94) = 3.50$, $p = .06$. This interaction indicates that the increase in difference between CSs paired with a positive versus a negative US from pre to post-acquisition was slightly more pronounced in the compound presentation group.

The analysis also revealed a main effect of time, $F(1, 94) = 17.39$, $p < .0001$, with more positive CS ratings before ($M = 2.01$, $SD = 17.62$) as compared to after acquisition ($M = -7.34$, $SD = 19.14$). Furthermore, a main effect of US-type was obtained, $F(1, 94) = 267.18$, $p < .0001$. This reflects the fact that across both experimental groups and time, CSs paired with liked USs (*M* = 26.15, *SD* = 22.58) were rated more positively than CSs paired with negative USs (*M* = -31.48, *SD* = 22.97). No other significant effects were obtained.

We can thus conclude that the conditioning phase was effective in inducing evaluative shifts in the expected direction in both experimental groups. The results further suggest that evaluative learning as indexed by ratings was slightly more pronounced in the compound presentation group as compared to the separate presentation group.

Priming data

A 2 (Presentation: compound vs. separate) x 2 (Time: pre vs. post-conditioning) x 2 (Block: 1 vs. 2) x 2 (Target: positive vs. negative) x 2 (US-type: intended positive vs. negative) ANOVA with repeated measures on the last four variables was performed. Importantly, a significant Time x Target x US-type interaction was obtained, $F(1, 92) = 19.19$, $p < .0001$. Planned comparisons revealed that whereas a nonsignificant Target x US-type interaction was obtained before EC took place, $F(1, 92) = 1.05$, *ns*, a highly significant interaction was observed after the EC phase, $F(1, 92) = 42.20$, $p < .0001$. Further planned comparisons on the post-test data revealed significantly shorter response latencies for positive (*M* = 605 ms, *SD* = 103) as compared to negative targets (*M* = 638, *SD* = 94) for the CSs that were paired with a liked US, $F(1, 92) = 24.60$, $p < .0001$. For the CSs that were paired with a disliked US, on the other hand, significantly shorter response latencies were observed for negative (*M* = 613, *SD* = 92) as compared to positive targets (*M* = 638, *SD* = 97), $F(1, 92) = 15.90$, $p < .001$. Hence, differential EC was obtained in the priming data. Of further importance is the absence of a four-way interaction between presentation, time, target and US-type, $F < 1$, which indicates that this differential EC effect did not differ between both experimental groups (see Table 1b for cell means). The results of the priming data thus corroborate the findings of the ratings.

The analysis further revealed two significant main effects. First, a significant main effect of time was observed, $F(1, 92) = 10.47$, $p < .01$, with faster responses after (*M* = 624, *SD* = 89) as compared to before conditioning (*M* = 641, *SD* = 83). There was also a significant main effect of block, $F(1, 92) =$

169.47, $p < .0001$. Significantly faster responses were observed in the second ($M = 610$, $SD = 85$) as compared to the first block ($M = 654$, $SD = 83$). In addition to these main effects, there was a significant Target x US-type interaction, $F(1, 92) = 33.1$, $p < .0001$. For CSs that were paired with a positive US, faster responses were observed for positive as compared to negative targets while the reverse was found for CSs paired with a negative US. This interaction was further qualified by a significant three-way interaction between presentation, target and US-type, $F(1, 92) = 5.41$, $p < .05$. This interaction indicates a stronger Target x US-type interaction (across time and block) in the compound presentation group as compared to the separate presentation group. Finally, a significant Time x Block interaction was obtained, $F(1, 92) = 15.46$, $p < .001$. A larger difference in response latencies was observed between the blocks of the pre-measurement as compared to the blocks of the post-measurement. No other significant effects were obtained.

Discussion

In the present experiment, the cookie-paradigm described by Verhulst et al. (2006) was compared with an adapted version in which the CS and US no longer formed a compound. The main findings indicate that differential EC was successfully demonstrated in both procedures. The results of the valence ratings indicated that previously neutral CSs became significantly more liked if paired with a positive flavor and significantly more disliked if paired with a negative flavor. Furthermore, significant differential EC was also obtained in the priming data.

The majority of participants in both experimental groups were classified as ‘contingency aware’. Moreover, since we used a post-conditioning assessment of contingency knowledge, participants’ level of contingency awareness *during* the experiment might even be underestimated due to forgetting. This raises the possibility that the EC effects found in the ratings could be due to demand characteristics. Importantly, the present results were not only obtained with a direct measure but also with an indirect measure, the APT, which is much less prone to demand bias. The present experiment therefore corroborates previous studies that found the APT capable of indirectly measuring (recently acquired) food attitudes (e.g., Lamote et al., 2004; Verhulst et al., 2006). Moreover, the current data extend these results by showing that the APT can be used as a pre/post assessment tool to measure *changes* in food attitudes over time.

As indicated earlier, for both procedures, the conditioning phase was effective in inducing evaluative shifts in the expected direction. Nevertheless, the marginally significant interaction between presentation, time and US-type found in the ratings suggests that evaluative learning was slightly more pronounced in the compound presentation group. As the proportion of aware participants did not differ between both experimental groups, this difference in evaluative learning cannot be due to a difference in contingency awareness between both groups. A possible explanation might then be found in the 'fit' between CS and US. Research indeed suggests that developing an associative CS-US link is easier when they show some sort of 'belongingness' or 'fit' (Hamm et al., 1989). Perhaps the closer CS-US connection in the compound presentation group resulted in larger conditioning effects. It can be argued, however, that a logical CS-US connection was also present in the separate presentation group as pictures of cookies (in contrast to, for instance, pictures of non-food items) were used as CSs. Moreover, the stronger EC effect in the compound presentation group was not replicated in the APT data. Hence, more research - in which the fit between CS and US is addressed more directly - is needed to clarify the possible role of 'belongingness' in evaluative learning.

Unexpectedly, a significant difference in evaluative ratings was observed between CSs that would be paired with a positive versus a negative US *before* EC took place. As CS-US assignment was counterbalanced over participants and the food stimuli were not visible before the start of the EC phase, we have no clear explanation for this finding. Moreover, this effect did not occur in the priming data. Importantly, even despite this pre-existing difference, clear evidence for successful evaluative learning was obtained in the ratings as indicated by the significant Time x US-type interaction. This finding nevertheless highlights the importance of including a pre-test in EC studies.

In the present experiment, sweet tastes were used as positive USs. Despite the fact that humans in general possess an innate preference for sweetness, attempts to demonstrate increased liking for CSs paired with sweet tastes have had mixed results (e.g., Baeyens, Eelen, Van den Bergh, & Crombez, 1990; Zellner, Rozin, Aron, & Kulish, 1983). Recent research suggests that people vary in the degree to which they experience sweet tastes as pleasant, with an apparent distinction between sweet likers and dislikers (Yeomans, Mobini, Elliman, Walker, & Stevenson, 2006). Studies which failed to obtain EC effects with a sweet US did not ensure or find that their participants perceived this US as positive. When participants were pre-selected on liking for sweetness (e.g., Yeomans, Leitch, Gould, & Mobini, 2008) or divided into sweet (dis)likers post-hoc (e.g., Baeyens et al., 1990) reliable EC effects were observed for sweet liking participants. These findings confirm that a prerequisite for EC is that the USs are experienced as intended and underline the importance of obtaining US ratings as a manipulation check or tool for pre-selection. In the present study US ratings were included and revealed that the vast majority of participants rated the USs as intended.

Further, both of the current cookie procedures can be conceptualized as visual-gustatory paradigms[17] where visual characteristics (i.e., color and shape) are paired with a gustatory US. Also here, some previous studies applying a color-taste paradigm failed to obtain EC effects (e.g., Baeyens et al., 1990). Further research is needed to investigate whether specific characteristics of the CSs used in this study are responsible for this difference in results. Possible factors that may play a role are, for instance, CS familiarity (e.g., Cacioppo, Marshall-Goodell, Tassinary, & Petty, 1992), CS-US belongingness or CS complexity. Furthermore, also differences in the contingency awareness of the

[17] Note that the present manuscript departs from a broad definition of 'food(preference)' in which not only the flavor but also the appearance of a food item is comprised. Both of the applied cookie procedures can be conceptualized as visual-gustatory paradigms that provide information on how we get to (dis)like the *appearance* or *packaging* of foods we eat. As several studies provide evidence that the appearance of a food influences whether it will be selected or ingested (e.g., Rolls, Rowe, & Rolls, 1982; Verhulst et al., 2006), we consider this aspect an important component of food preference and apply the term 'food preference learning' to refer to learning about the visual aspects of a food item throughout the manuscript. Thus, it is important to keep in mind that in the present manuscript 'food preference learning' is broader than 'flavor preference learning'. The appearance of a food item might also influence how its flavor is perceived, but this remains to be investigated.

participants or the strength of the applied USs might have contributed and need further study.

Most importantly, our adapted cookie-paradigm resulted in strong EC effects, as evidenced by both ratings and the APT. Therefore, we conclude that our adapted paradigm constitutes a promising tool for studying conditioning processes in food preference learning. An important advantage of this paradigm is that it allows for separate CS or US presentations and therefore is suitable for studying more complex conditioning processes like extinction or latent inhibition. For instance, as liking is an important determinant of food choice and intake, an interesting question for health practitioners is whether acquired unhealthy food preferences (and eating habits based on these) can be changed through an extinction procedure. The present paradigm facilitates future research examining this and other issues related to the role of complex conditioning processes in food preference learning. Therefore, we believe our adapted procedure constitutes a convenient tool to learn more about how food preferences are formed and changed. In addition, the fact that strong EC effects were obtained - both in the ratings and the APT - with this paradigm also makes it a good candidate for studying some possible boundary conditions in EC. Some studies, for instance, suggest that flavor-based evaluative learning might be attenuated when participants are restrained eaters or sated (e.g., Brunstrom, Downes, & Higgs, 2001). Future research could investigate the influence of such factors on evaluative learning in the present cookie-paradigm.

References

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation, 21*(4), 434-455.

Brunstrom, J. M., Downes, C. R., & Higgs, S. (2001). Effects of dietary restraint on flavour-flavour learning. *Appetite, 37*(3), 197-206.

Cacioppo, J. T., Marshall-Goodell, B., Tassinary, L., & Petty, R. E. (1992). Rudimentary determinants of attitudes - Classical-conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology, 28*(3), 207-233.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853-869.

Eertmans, A., Baeyens, F., & Van den Bergh, O. (2001). Food likes and their relative importance in human eating behavior: Review and preliminary suggestions for health promotion. *Health Education Research, 16*(4), 443-456.

Hamm, A. O., Vaitl, D., & Lang, P. J. (1989). Fear conditioning, meaning, and belongingness - A selective association analysis. *Journal of Abnormal Psychology, 98*(4), 395-406.

Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology, 52*(3), 180-186.

Hermans, D., Clarysse, J., Baeyens, F., & Spruyt, A. (2005). Affect (Version 4.0) [Computer software; retrieved from http://www.psy.kuleuven.ac.be/leerpsy/affect4]. University of Leuven, Belgium.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34,* 115-139.

Lamote, S., Hermans, D., Baeyens, F., & Eelen, P. (2004). An exploration of affective priming as an indirect measure of food attitudes. *Appetite, 42*(3), 279-286.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology - Learning Memory and Cognition, 33*(1), 130-144.

Rolls, B. J., Rowe, E. A., & Rolls, E. T. (1982). How flavor and appearance affect human feeding. *Proceedings of the Nutrition Society, 41*(2), 109-117.

Rozin, P., & Millman, L. (1987). Family environment, not heredity, accounts for family resemblances in food preferences and attitudes - A twin study. *Appetite, 8*(2), 125-134.

Verhulst, F. (2007). *The affective priming paradigm as an indirect measure of food attitudes*. Unpublished doctoral dissertation, University of Leuven, Leuven.

Verhulst, F., Hermans, D., Baeyens, F., Spruyt, A., & Eelen, P. (2006). Determinants and predictive validity of direct and indirect measures of recently acquired food attitudes. *Appetite, 46*(2), 137-143.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(4), 454-459.

Yeomans, M. R., Leitch, M., Gould, N. J., & Mobini, S. (2008). Differential hedonic, sensory and behavioral changes associated with flavor-nutrient and flavor-flavor Learning. *Physiology & Behavior, 93*(4-5), 798-806.

Yeomans, M. R., Mobini, S., Elliman, T., Walker, H. C., & Stevenson, R. J. (2006). Hedonic and sensory characteristics of odors conditioned by pairing with tastants in humans. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(3), 215-228.

Zellner, D. A., Rozin, P., Aron, M., & Kulish, C. (1983). Conditioned enhancement of humans liking for flavor by pairing with sweetness. *Learning and Motivation, 14*(3), 338-350.

experiment 3

Counterconditioning: An effective technique for changing conditioned preferences

The evidence for the effectiveness of counterconditioning as a strategy for changing conditioned preferences is rather scarce and inconclusive. The present experiment reinvestigated this issue and compared the effect of further conditioning, extinction and a counterconditioning procedure on recently acquired conditioned preferences in a picture-taste paradigm. Self-report and affective priming data indicated that whereas further conditioning and extinction trials were ineffective in fully eliminating the previously acquired evaluations, the counterconditioning treatment did succeed in doing this. A follow-up valence assessment revealed that all these effects persisted after a 7-day delay period.

In press as: Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (in press). Counterconditioning: An effective technique for changing conditioned preferences. *Experimental Psychology.*

Introduction

Preferences influence virtually all human behavior; they impact upon our social interactions (e.g., Walther, Nagengast, & Trasselli, 2005), the foods we eat (e.g., Eertmans, Baeyens, & Van den Bergh, 2001), the products we buy (e.g., Gibson, 2008), our emotions (e.g., Hermans, De Houwer, & Eelen, 1996; Sherer, 1993), etc. Therefore, understanding how (dis)likes are formed and subsequently changed is crucial.

An important route for preference acquisition is evaluative conditioning (EC), which refers to changes in the evaluation of a stimulus that result from pairings with other positive or negative stimuli (see De Houwer, Baeyens, & Field, 2005; De Houwer, Thomas, & Baeyens, 2001 for extensive reviews on this topic). In a prototypical EC experiment, a neutral stimulus (Conditioned Stimulus; CS) is presented repeatedly with a subjectively liked or disliked stimulus (Unconditioned Stimulus; US). After repeated pairings, the CS valence typically shifts in the direction of the US valence. Preference learning through EC has been repeatedly demonstrated for a variety of stimuli (De Houwer et al., 2001). The EC paradigm therefore provides an interesting framework for experimentally investigating preference formation processes.

Whereas past studies predominantly focused on acquisition/formation processes, recently, interest has developed in how preferences, once acquired, can be changed. Remarkably, although EC resembles Pavlovian conditioning procedurally, it differs in being less susceptible to extinction (e.g., Diaz, Ruiz, & Baeyens, 2005; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006; but see Lipp, Oughton, & Lelievre, 2003).

Importantly, this lower extinction sensitivity does not make conditioned changes in liking unchangeable. The literature reports two possible procedures for altering conditioned preferences: US-revaluation and counterconditioning. The former entails that a conditioned change in liking can be reversed by altering the valence of the original US. Baeyens, Eelen, Van den Bergh, and Crombez (1992), and more recently Walther, Gawronski, Blank, and Langer (2009), indeed demonstrated that post-conditional changes in the valence of a US cause corresponding changes in the valence of pre-associated CSs.

The second procedure, counterconditioning, implies pairing the CS with a US having a valence opposite from the original US. Although this procedure has been well studied in animals (Dickinson & Pearce, 1977), only few studies have focused on counterconditioning - and, more specifically, its impact on evaluative learning - in human beings. Baeyens, Eelen, Van den Bergh, and Crombez (1989) were the first to observe that the conditioned positive/negative valence of a CS can be eliminated or even reversed by pairing it with a new negative/positive US. This first study is, however, considered inconclusive by some authors because CS-US pairs were not randomized but arranged by the experimenter on the basis of perceptual similarity. Field and Davey (1999) demonstrated that such a CS-US assignment can result in artifactual EC effects. They obtained conditioning-like changes in the evaluation of the CSs when the CS-US pairs were constructed like this even in those participants who were never exposed to actual CS-US pairings. Therefore, the observed changes in valence in Baeyens et al.'s (1989) study might be due to similarity effects rather than associative learning. Furthermore, also demand effects cannot be ruled out as valence was only measured by self-report. Nevertheless, surprisingly little follow-up research has been conducted on counterconditioning as a strategy for changing conditioned preferences. Only Stevenson, Boakes, and Wilson (2000) further investigated this issue and failed to replicate the results of Baeyens et al. (1989). Using an odor-taste paradigm, Stevenson et al. (2000) observed no difference in liking between an odor mixed with citric acid only and an odor first mixed with citric acid and subsequently with sucrose. This replication failure might, however, relate to their choice of US. Possibly their intended positive US (i.e., sucrose) was not perceived as such by all participants as the authors also observed no increased liking for an odor consistently paired with sucrose.

In sum, the evidence for the effectiveness of counterconditioning as a strategy for changing conditioned preferences seems rather scarce and inconclusive. Therefore, we aimed at reinvestigating this issue. Note that recently some studies have focused on (partially) related matters, like the impact of evaluative counterconditioning on *long existing* attitudes (e.g., self-esteem, Dijksterhuis, 2004; ageism, Karpinski & Hilton, 2001; prejudice, Olson & Fazio, 2006). In these studies it is unclear, however, how and when participants acquired their ‘a-priori’ preferences. Therefore, these reports provide little information on whether counterconditioning operates similarly for preferences (recently) acquired through EC. Finally, some social psychology studies have investigated the impact of counter-attitudinal information on recently acquired attitudes using an *impression formation paradigm*. In this procedure, participants received information about whether certain positive and negative behaviors were (un)characteristic of a fictional person. In several experiments, Rydell and colleagues (Rydell & McConnell, 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007) observed that participants’ experimentally induced attitudes could be changed with additional contradicting information. Nevertheless, it remains unclear whether a (more typical) EC procedure would yield similar results.

The present study investigated evaluative counterconditioning using a picture-taste paradigm that previously resulted in strong evaluative learning effects (Kerkhof, Vansteenwegen, Baeyens, & Hermans, 2009). To control for similarity effects, CS-US pairs were constructed randomly. To control for demand biases, an indirect valence measure (an affective priming task, APT) was used in addition to ratings.

The experiment consisted of two phases. During the acquisition phase, three CS-pictures were paired with a positive US while three other CSs were combined with a negative US. In the post-acquisition phase, two CSs were presented with the same US as in the first phase (*further conditioning pair*), two were presented without US (*extinction pair*), and two were presented with a US of opposite value (*counterconditioning pair*). Valence was assessed before acquisition and after the second conditioning phase by ratings and an APT. Based on previous studies, we expected the CSs of the ‘extinction pair’ to acquire and maintain the valence of their acquisition US. For the CSs of the ‘counterconditioning pair’, on the other hand, we expected to observe a neutral value or one in the direction of the post-acquisition US. Baeyens, Eelen, Crombez, and Van den Bergh (1992) observed that additional CS-US pairings do not always strengthen the EC effect, but may even weaken it. Therefore the ‘further conditioning pair’ was included to see whether counterconditioning results in stronger changes than additional conditioning trials.

Finally, we explored whether counterconditioned preferences remain intact or regress to their initial meaning (i.e., spontaneous recovery, Bouton, 2004) over time by reassessing valence at 1-week follow-up.

Method

Participants

Sixty-three first-year psychology students (55 women) participated.

Materials

Six pictures (512 x 384 pixels) of cookies, differing in color and shape, served as CSs. They were a brown triangle, red moon, green star, dark-yellow pear shape, fluo-yellow diamond, and white three-leaf clover shape.

Good and bad tasting cookies served as USs. The pastry was made with 200 g flour, 100 g sugar ('good' cookies only), one egg, and 125 g butter. It was divided in two halves and by adding two tablespoons of hazelnut or four tablespoons of Tween20 (*polysorbate 20*; a bitter soapy tasting chemical) 'intended' good, respectively bad, tasting cookies were obtained. All cookies were yellow squares with a side length of 1 cm. CS-US pairings were counterbalanced over participants using a Latin square scheme.

For the APT, the CS-pictures served as primes, and 10 positive and 10 negative Dutch nouns (selected from Hermans & De Houwer, 1994) were used as targets (for more details on the APT, see the Procedure section below).

An Affect 4.0 program (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010) controlled stimulus presentation and response registration.

Procedure

Participants were tested in a laboratory containing 10 identical separated test boots, arranged in a semicircle, such that visual contact between participants was precluded, but each participant could clearly see and hear the experimenter. The experimenter was seated behind a table at the center of the semicircle. The experiment consisted of two sessions.

Session 1. This session started with a pre-acquisition measurement of the valence of each CS on scale ranging from -100 (*very unpleasant)* to +100 (*very* *pleasant*) in steps of 10. Valence was also assessed by means of an APT. The experimenter explained that on each trial a picture (called the ‘prime’) would precede a word (called the ‘target’). Participants were instructed to attend the word and evaluate it as quickly as possible by pressing the right, respectively left, mouse button for positive, respectively negative, words. The priming task consisted of two blocks of 60 experimental trials, preceded by 12 practice trials. Each trial started with a fixation cross (500 ms), followed by a 500 ms blank screen, followed by the prime (200 ms). Fifty milli seconds after prime offset, the target was presented until participants responded or 2000 ms elapsed. The mean intertrial interval was 1000 ms (range 500-1500 ms). Within each block every prime was paired with five negative and five positive targets. Each target appeared equally often in each block. Task order (priming-rating vs. rating-priming) was counterbalanced across participants.

During the conditioning phases, participants remained seated in front of the computer monitor where the CS-pictures were presented one by one. Some pictures were followed by the presentation of the message ‘Take and eat a cookie’ while no message was presented after others. Participants were further presented with a cup of water and a piece of plastified paper on which 72 small cookies were arranged in six rows of 12. At the beginning of each row was a small picture of one of the CSs. Participants were asked to attend the pictures on the computer screen and were instructed to eat a cookie from the row containing this picture *only* when it was followed by the message ‘Take and eat a cookie’. To avoid aftertastes, participants were asked to drink some water between trials. Thereafter, they were instructed to press the ‘Enter’ key to proceed to the next trial. An overview of the sequence and timing of the different events within a conditioning trial is given in Table 2.

The acquisition phase consisted of 36 trials with six presentations per CS. The message ‘Take and eat a cookie’ was presented 1000 ms after the onset of *each* CS. Three of the six CSs were consistently paired with a positive US, while the others were paired with a negative US. Presentation order was randomized for every participant, with the restriction that no CS-picture appeared twice in a row. A one-minute break separated the two conditioning phases.

Table 2
Sequence and timing of events within a conditioning trial

Time within trial	Event(s)
0 ms	Onset CS
1000 ms	Onset message 'Take and eat a cookie'
	Participants take and eat a cookie
21000 ms	Participants are instructed to take a sip of water
41000 ms	Participants are instructed to press ENTER to proceed to the next trial
41500 ms	Intertrial interval (black screen)

Note. The message 'Take and eat a cookie' was not presented during extinction trials. During these trials, participants waited until they were instructed to press the enter key to proceed to the following trial.

The post-acquisition phase also consisted of 36 trials, with six presentations per CS. During this phase two CSs (one previously paired with a positive US, and one previously combined with a negative US) received further conditioning trials with the same US as during acquisition. Two other CSs were presented without US (i.e., were not followed by 'Take and eat a cookie'). The last two CSs were followed by a US of opposite valence (counterconditioning trials). All other presentation parameters were identical to those of the acquisition phase.

After the post-acquisition phase participants again provided pleasantness ratings for the cookie-pictures and completed a second APT.

Next, contingency recall was assessed by asking participants to indicate for each of the six CSs for both conditioning phases separately whether it had been followed by a hazelnut cookie, a Tween20 cookie, or no cookie and how certain they were of this *(very uncertain, reasonably uncertain, reasonably certain,* and *very certain)*. Finally, participants rated the US flavors on a scale ranging from -100 *(very unpleasant)* to +100 *(very pleasant)*.

Session 2. One week later, participants provided pleasantness ratings and completed the APT for a third time. Contingency recall was also assessed and participants were asked to rate the US flavors one last time.

Results

Preliminary analyses

Eight participants who failed to comply with task instructions were excluded from the priming analyses. The data were corrected for outliers by excluding reaction times (RTs) below 200 ms or above 1500 ms (0.64%). Trials with no (0.26%) or incorrect responses (7.83%) were also discarded. To create an evaluation score, the mean latency for positive target words was subtracted from the mean latency for negative words for each of the six prime categories (Gawronski, Walther, & Blank, 2005). Higher scores on this variable indicate more positive evaluations. Note that these scores should not be interpreted in an absolute manner (e.g., a value of zero reflecting a neutral evaluation), because response latencies for positive target words may generally differ from response latencies for negative target words.

US ratings

The US ratings were analyzed using a 2 (US-type: intended positive, negative) x 2 (Moment: after conditioning, after a week) ANOVA. A main effect of US-type was obtained, $F(1, 62) = 791.03$, $p < .0001$, indicating that the intended positive US was rated significantly more positive ($M = 63.25$, $SD = 28.98$) than the intended negative US ($M = -83.10$, $SD = 21.62$). The main effect of moment and the Moment x US-type interaction were not significant, both $F < 1$. Hence, the USs were rated as intended both following the experiment and one week later.

CS ratings

A 2 (Acquisition valence: positive, negative US) x 3 (Post-acquisition contingency: counterconditioning, extinction, further conditioning) x 3 (Moment: before conditioning, after conditioning, after a week) repeated measures ANOVA was conducted. The crucial three-way interaction between acquisition valence, post-acquisition contingency, and moment was significant, $F(4, 248) = 30.78$, p

< .0001, and further examined with separate Acquisition valence x Post-acquisition contingency analyses for each test moment.[18]

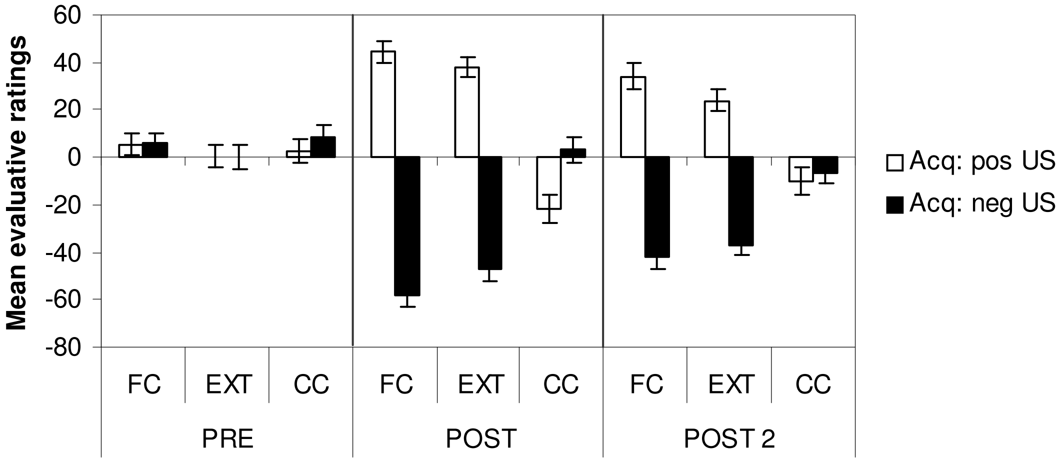


Figure 3. Mean evaluative ratings of the CSs that received acquisition trials with a positive (Acq: pos US) or negative US (Acq: neg US) and subsequent further conditioning (FC), extinction (EXT) or counterconditioning (CC) presentations as a function of moment (PRE: before conditioning, POST: after conditioning, POST 2: after a week). Error bars represent standard errors.

Moment 1: before conditioning (Figure 3, left panel). We expected no differences in the valence ratings of the CSs before conditioning. Consistent with this prediction, no significant main or interaction effects appeared at baseline, all Fs < 1.

Moment 2: after conditioning (Figure 3, middle panel). After the first session, we expected the CSs that were involved in extinction or further conditioning trials to have (at least partially) maintained the valence of their acquisition US. In contrast, for the CSs that underwent counterconditioning trials, we expected a neutral value or a value in the direction of the post-acquisition US.

A significant main effect of acquisition valence, $F(1, 62) = 97.65$, $p < .0001$, and a significant Acquisition valence x Post-acquisition contingency interaction, $F(2, 124) = 92.43$, $p < .0001$, emerged. Planned comparisons revealed a significant EC effect (i.e., main effect of acquisition valence) at the

[18] The analysis further revealed main effects of moment and acquisition valence. Additionally, there was a moment by acquisition valence interaction and an acquisition valence by post-acquisition contingency interaction. For reasons of brevity, these effects are not discussed. A complete set of analyses is available from the authors.

'further conditioning' and 'extinction' levels of the post-acquisition contingency factor, with $F(1, 62) = 185.17$, $p < .0001$, and $F(1, 62) = 159.55$, $p < .0001$, respectively. As expected, after the post-acquisition phase, the CSs that were paired with a positive US during acquisition were still rated significantly more positive ($M_{fc} = 44.62$, $SD_{fc} = 35.85$, $M_{ext} = 38.41$, $SD_{ext} = 33.47$) than those paired with a negative acquisition US ($M_{fc} = -58.41$, $SD_{fc} = 38.66$, $M_{ext} = -47.46$, $SD_{ext} = 36.28$). Both treatments failed to fully eliminate the evaluative value that the CSs obtained during acquisition. In contrast, the influence of the acquisition contingencies was abolished and even reversed by the counterconditioning treatment, as indicated by a reversed EC effect at the 'counterconditioning' level of the post-acquisition contingency factor, $F(1, 62) = 6.78$, $p < .05$. The counterconditioning treatment caused the CSs to acquire an evaluative value of the opposite sign and resulted in a more positive evaluation of the CS that was paired with a negative US during acquisition and a positive US during counterconditioning ($M = 3.33$, $SD = 42.627$) compared to the CS that was first paired with a positive US and subsequently with a negative US ($M = -21.75$, $SD = 48.78$).

A further set of contrasts comparing the post-acquisition treatments confirmed these results and revealed a significantly stronger EC effect for the further conditioning and extinction treatments, compared to the counterconditioning treatment, with $F(1, 62) = 112.59$, $p < .0001$, and $F(1, 62) = 98.30$, $p < .0001$. Additionally, a stronger EC effect was observed in the further conditioning condition as compared to the extinction condition, $F(1, 62) = 6.86$, $p < .05$.

Based on the rating data collected after conditioning, we can conclude that whereas the value that the CSs acquired during acquisition was not fully eliminated by the further conditioning or by the extinction treatment, it was abolished and even slightly reversed by the counterconditioning treatment.

Moment 3: after a week (Figure 3, right panel). A similar pattern of results emerged one week later. Also at follow-up, a significant main effect of acquisition valence, $F(1, 62) = 72.13$, $p < .0001$, and a significant Acquisition valence x Post-acquisition contingency interaction, $F(2, 124) = 35.16$, $p < .0001$, were obtained. Planned comparisons again revealed a significant EC effect at the 'further conditioning' and 'extinction' levels of the post-acquisition contingency factor, with $F(1, 62) = 69.07$, $p < .0001$, and $F(1, 62) = 82.57$, $p < .0001$. More positive ratings were still obtained for CSs that had been paired with a positive acquisition US ($M_{fc} = 33.97$, $SD_{fc} = 44.38$, $M_{ext} = 24.13$, $SD_{ext} = 35.81$) compared to a negative acquisition US ($M_{fc} = -42.22$, $SD_{fc} = 40.78$, $M_{ext} = -36.83$, $SD_{ext} = 33.01$). Thus, the evaluative value acquired during the first session (at least partially) persisted over a 1-week delay. No effect of acquisition valence emerged at the 'counterconditioning' level, $F < 1$. Hence, after one week the influence of the initial acquisition contingencies was still abolished, but no longer reversed.

Additional contrasts comparing the post-acquisition treatments revealed that there was still a significantly stronger EC effect for the further conditioning and extinction treatments compared to the counterconditioning treatment, with $F(1, 62) = 48.30$, $p < .0001$, and $F(1, 62) = 43.09$, $p < .0001$. After a week, however, the further conditioning and extinction treatments no longer differed, $F(1, 62) = 2.99$, *ns*.

Based on the ratings collected in Session 2, we can conclude that the value that the CSs acquired during the first session (at least partially) persisted after a week. Importantly, the effect of the counterconditioning treatment also persisted as for the CSs that underwent this treatment no influence of the association that was made during the acquisition phase could be detected.

Priming data

To test for similar effects in the priming data, a 2 (Acquisition valence) x 3 (Post-acquisition contingency) x 2 (Block: 1, 2) x 3 (Moment) repeated measures ANOVA was performed on these data. The crucial three-way interaction between acquisition valence, post-acquisition contingency and moment was significant, $F(4, 216) = 3.40$, $p < .05$, and subjected to separate analyses for each test moment.[19]

[19] The analysis also revealed main effects of moment and acquisition valence. In addition, there was a Moment x Acquisition valence interaction and an Acquisition valence x Post-acquisition contingency interaction. Again, for brevity, these effects are not discussed.

<u>*Moment 1: before conditioning (Figure 4, left panel).*</u> No differences in the evaluation of the three CS-pairs (further conditioning, extinction, and counterconditioning) were expected before conditioning. In line with this expectation, the analysis revealed a nonsignificant main effect of acquisition valence and a nonsignificant Acquisition valence x Post-acquisition contingency interaction, both F's < 1.

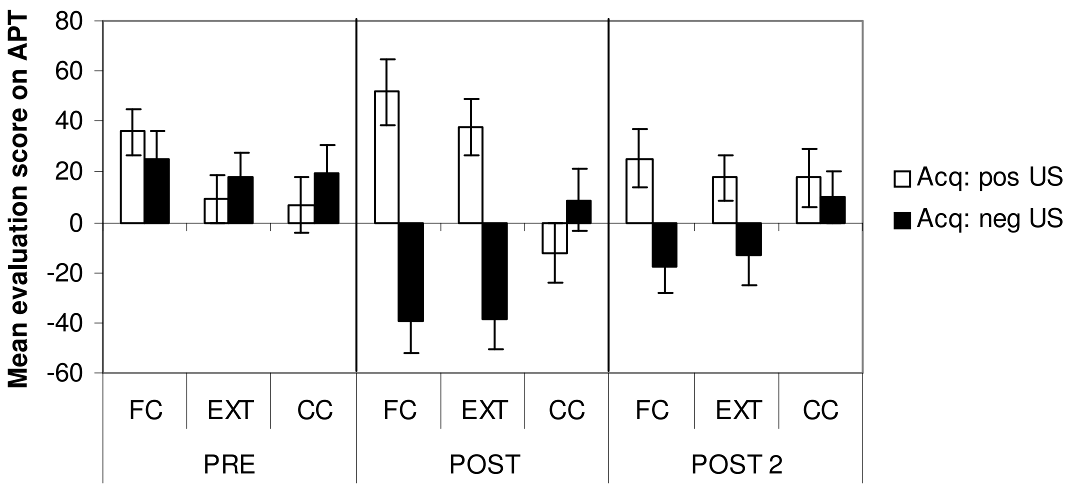


Figure 4. Mean evaluation score on the APT for CSs that received acquisition trials with a positive (Acq: pos US) or negative US (Acq: neg US) *and* subsequent further conditioning (FC), extinction (EXT) or counterconditioning (CC) presentations as a function of moment (PRE: before conditioning, POST: after conditioning, POST 2: after a week). Error bars represent standard errors.

<u>*Moment 2: after conditioning (Figure 4, middle panel).*</u> After the first session, we expected the CSs that underwent extinction or further conditioning trials to have (at least partially) maintained the valence of their acquisition US. Conversely, for the CSs that underwent counterconditioning trials we expected a neutral value or a value in the direction of the post-acquisition US. Regarding the priming data, this implies the prediction of a significant effect of acquisition valence (i.e., EC effect) for the extinction and further conditioning CS-pairs and a nonsignificant or reversed effect of acquisition valence for the counterconditioning pair.

The analysis revealed a significant main effect of acquisition valence, $F(1, 54) = 20.67$, $p < .0001$, and a significant Acquisition valence x Post-acquisition contingency interaction, $F(2, 108) = 10.26$, $p < .0001$. Planned comparisons confirmed a significant EC effect (i.e., acquisition valence effect) at the 'further conditioning' and 'extinction' levels of the post-acquisition contingency factor, with $F(1, 54) = 22.34$, $p < .0001$, and $F(1, 54) = 18.90$, $p < .0001$. As expected, more positive scores were obtained for the CSs that were paired with a positive US during acquisition ($M_{fc} = 52$, $SD_{fc} = 96$, $M_{ext} = 38$, $SD_{ext} = 83$) as compared to the CSs that were combined with a negative US during acquisition ($M_{fc} = -39$, $SD_{fc} = 96$, $M_{ext} = -38$, $SD_{ext} = 92$). These results corroborate the findings of the ratings in suggesting that the influence of the acquisition contingencies was not fully abolished by the further conditioning and extinction treatments. In contrast, at the 'counterconditioning' level, no significant effect of acquisition valence appeared, $F(1, 54) = 1.11$, *ns*. There was even a trend for a more positive evaluation of the CS that was paired with a negative US during acquisition and a positive US during counterconditioning ($M = 9$, $SD = 92$) compared to the CS that was first paired with a positive US and subsequently with a negative one ($M = -12$, $SD = 87$). This again corroborates the findings of the ratings and suggests that the influence of the initial acquisition contingencies was eliminated (and even slightly reversed) by the counterconditioning treatment.

A further set of contrasts comparing the post-acquisition treatments confirmed these results and revealed a significantly larger effect of acquisition valence for the further conditioning and extinction treatments compared to the counterconditioning treatment, with $F(1, 54) = 12.35$, $p < .001$, and $F(1, 54) = 13.74$, $p < .001$. The effect of acquisition valence did not differ between the further conditioning and extinction treatments, $F < 1$.

In sum, the priming data collected after conditioning perfectly corroborate the ratings in suggesting that whereas the further conditioning and extinction trials failed to fully eliminate the valence the CSs acquired during the initial acquisition phase, the counterconditioning trials did succeed in this.

Moment 3: after a week (Figure 4, right panel). After a week, there was still a significant main effect of acquisition valence, $F(1, 54) = 11.42$, $p < .01$, but the Acquisition valence x Post-acquisition contingency interaction was no longer significant, $F(2, 108) = 1.41$, *ns.* Given our a-priori hypotheses, we nevertheless examined the effect of acquisition valence separately for the different post-acquisition treatments. These planned comparisons revealed a significant effect of acquisition valence at the ‘further conditioning’ and ‘extinction’ levels, with $F(1, 54) = 7.57$, $p < .01$, and $F(1, 54) = 4.47$, $p < .05$. More positive scores were observed for the CSs that were paired with a positive US during acquisition ($M_{fc} = 25$, $SD_{fc} = 86$, $M_{ext} = 17$, $SD_{ext} = 69$) than for the CSs that were combined with a negative US ($M_{fc} = -17$, $SD_{fc} = 78$, $M_{ext} = -13$, $SD_{ext} = 87$). Hence, these data corroborate the results of the ratings in suggesting that the influence of the acquisition contingencies (at least partially) persisted after a week. There was no effect of acquisition valence at the ‘counterconditioning’ level, $F < 1$, suggesting that the counterconditioning effect also persisted (with $M_{CSposneg} = 18$, $SD_{CSposneg} = 84$, $M_{CSnegpos} = 10$, $SD_{CSnegpos} = 76$). Again, this is in line with the ratings.

In conclusion, the priming data of Session 2 largely corroborate the rating results in suggesting that the evaluative value that the CSs acquired during the first session became a little weaker but persisted after a week. Importantly, the effect of the counterconditioning treatment also persisted as the effect of acquisition valence remained nonsignificant.

Contingency recall data

A participant was considered to remember a certain CS-US association when he/she correctly identified the US that had been paired with a specific CS and was at least ‘reasonably certain’ of his/her choice. The mean number of recalled pairings was 3.39 ($SD = 1.44$, range 0-6) for the acquisition contingencies and 4.14 ($SD = 1.47$, range 1-6) for the post-acquisition contingencies for Session 1, and 2.79 ($SD = 1.53$, range 0-6) for the acquisition contingencies and 2.54 ($SD = 1.64$, range 0-6) for the post-acquisition contingencies for Session 2, all out of a maximum of 6. A 2 (Moment: after conditioning, after a week) x 2 (Contingencies: acquisition, post-acquisition) ANOVA revealed a main effect of moment, $F(1, 62) = 44.81$, $p < .0001$, with fewer remembered contingencies after a week as compared to immediately after the experiment. Furthermore, there was a significant Moment x Contingencies interaction, $F(1, 62) = 13.56$, $p < .0001$. Whereas immediately after conditioning more post-acquisition contingencies were correctly identified as compared to acquisition contingencies, $F(1, 62) = 8.49$, $p < .01$, no such difference appeared after a week, $F < 1$.

Discussion

This experiment examined the effect of a further conditioning, extinction, and counterconditioning procedure on recently acquired conditioned preferences in a picture-taste paradigm. The main findings indicate that whereas further conditioning and extinction trials did not fully eliminate the valence the CSs acquired during acquisition, the counterconditioning treatment did. These effects appeared both in the ratings and APT and persisted over a 1-week delay.

The present results corroborate the findings of Baeyens et al. (1989), who also observed that evaluative learning was more sensitive to counterconditioning than to extinction. Important merits of the current study are the random CS-US assignment and the use of an indirect valence measure next to ratings. The findings concerning the more limited impact of extinction on conditioned valence are, furthermore, in line with several previous studies (e.g., Diaz et al., 2005; Vansteenwegen et al., 2006). The observation that evaluative learning is little or not sensitive to extinction has, among other findings (see De Houwer et al., 2001), led some authors to suggest that even though EC procedurally resembles other forms of Pavlovian conditioning, different processes may underlie it. Baeyens and colleagues (Baeyens & De Houwer, 1995; Baeyens, Eelen, & Crombez, 1995), for instance, have argued that EC is mediated by a mechanism that is qualitatively different from that which mediates Pavlovian preparatory conditioning. While the latter is seen as a form of *signal learning* where the CS generates the expectancy/belief that the US will occur in the here and now, Baeyens and co. (Baeyens & De Houwer, 1995; Baeyens et al., 1995) claim that the mechanism underlying EC is probably merely *referential*, in that the CS activates the ‘thought’ of the US without necessarily generating the expectancy that the US will actually occur. According to Baeyens and colleagues the difference in extinction susceptibility between EC and Pavlovian preparatory conditioning can then be explained by the fact that non-occurrence of the US disconfirms the predictive value of the CS but leaves its capacity to make one ‘think of’ the US intact (i.e., the CS retains its referential value).

Hence, whereas signal learning is sensitive to the statistical CS-US contingency, referential learning may be merely responsive to spatiotemporal contiguity (the co-occurrence of events). Our finding that evaluative learning is susceptible to counterconditioning is in line with this referential account, as counterconditioning implies the co-occurrence of a CS with a new US.

The contingency recall data suggest that participants, especially at the end of the first session, remembered most of the experimental contingencies. This raises the possibility that the effects found in the ratings (and especially those of Session 1) are due to demand characteristics. Importantly, the present results were not only obtained with ratings but also with an indirect measure, an APT, which is much less prone to demand bias.[20] First of all, the APT provides an indirect measure of stimulus valence. Participants are not directly asked to consciously reflect on their attitudes. Rather, their evaluation is inferred from their response towards an irrelevant and unrelated stimulus (the target), which makes this measure less transparent for (and therefore less controllable by) participants. Second, the presentation parameters that were used in the APT make the development of conscious response strategies unlikely. The short interval (i.e., 250 ms) between the onset of the prime and the onset of the target (known as the Stimulus Onset Asynchrony or SOA) makes it improbable that participants had the time to intentionally process and strategically use the valence of the prime (see Hermans, De Houwer, & Eelen, 2001; Hermans, Spruyt, & Eelen, 2003 for a more detailed discussion on this topic).

Although generally a trend was observed for stronger EC effects after further conditioning compared to extinction, a significant difference between both treatments emerged only in the ratings collected immediately after conditioning. This difference disappeared, moreover, when the baseline value of the CSs was taken into account: Contrasts comparing the evaluative ratings before conditioning with those after conditioning revealed no difference in the increase in differentiation between the CSs of the further conditioning and extinction

[20] Supporting the convergent validity of our direct and indirect measure, a regression analysis controlling for the covariate subject revealed a significant positive correlation between participants' ratings and priming scores, with r = .33, p < .001 after conditioning and r = .35, p < .001 after a week. Subject was added as a covariate to control the fact that each participant contributed multiple observations (i.e., for each participant six rating scores and six priming scores (one for each CS) were available).

treatments, $F(1, 62) = 2.49$, *ns*. Therefore, our data suggest that additional conditioning trials do not necessarily result in substantially stronger EC effects (see also Hofmann, De Houwer, Perugini, Baeyens, & Crombez, in press). Because valence was not assessed immediately after acquisition, the present design allows no further conclusions on the precise impact of both treatments. It is, for instance, unclear whether the further conditioning and extinction trials left the previously acquired valence of the CSs completely unchanged or resulted in a (small) reduction or increase in the EC effect obtained after acquisition. We chose not to include an additional test immediately after acquisition because repeated administrations of an indirect RT measure are known to result in reduced effects (e.g., Greenwald & Nosek, 2001). Moreover, as our main interest was to compare the effect of the different post-acquisition treatments, we wanted to maximize the power of our 'post-acquisition' test, and hence omitted further testing immediately after acquisition. Nevertheless, it would be interesting for future research to also examine the effect of these treatments separately (e.g., by including a measurement immediately after acquisition) as, for instance, despite its obvious relevance to EC, relatively little is known about the (mechanism behind the) influence of the number of CS-US pairings on EC effects (see also De Houwer et al., 2001).

The data obtained during Session 2 clearly indicated that the effects observed immediately after conditioning persisted over a week. The slightly weaker priming effects observed in this session possibly relate to the above-mentioned problem that prior experience with an indirect RT task can lead to reduced effects. Also, floor-effects might have been at play due to the reoccurrence of the same target words in the different priming tasks. Participants indeed responded considerably faster during the last priming task as compared to the previous two. Hence, it might be interesting for future research to use different sets of target words when a priming task is repeatedly administered. Another alternative would be using an online variant of the APT (Kerkhof, Goesaert, et al., 2009), which has the additional advantage of providing more insight into the changes in evaluation of the CSs throughout the experiment.

Besides theoretical importance, the present findings potentially have important implications for preference change in applied settings. The current findings, for instance, suggest that counterconditioning forms a good alternative to US-revaluation for changing conditioned preferences. An advantage of this procedure, moreover, is that knowledge of the initial acquisition US is not required. Investigating the impact of counterconditioning on pre-existing preferences with an unknown acquisition history therefore becomes possible (e.g., Dijksterhuis, 2004; Karpinski & Hilton, 2001; Olson & Fazio, 2006). An example of the latter can be found in clinical fears. Fear conditioning research suggests that signal and referential learning can co-occur in clinical fears that are based on the experience of a contingency between an originally neutral stimulus (e.g., an elevator) and a threatening stimulus (e.g., a panic attack). Throughout conditioning the CS not only becomes a predictor for the negative experience, but also gains negative valence because of its association with the aversive US. The findings concerning the weak extinction susceptibility of evaluative learning suggest that a standard exposure intervention (the clinical analogue of extinction), might successfully reduce the expectancy component in clinical fear, but might leave the acquired negative meaning of the phobic object relatively unaltered. As there is evidence that this remaining negative valence could form an affective-motivational source for the re-emergence of the original phobic fear (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004), it might be therapeutically beneficial to combine exposure with techniques (like counterconditioning or US-revaluation) aimed at altering the valence of the fear-eliciting stimulus. A common observation is that many phobics do not recall having had a traumatic experience in the presence of their phobic object (Mineka & Zinbarg, 2006). In such cases it is unclear how US-revaluation could be applied to reduce the phobic object’s acquired negative valence, whereas a counterconditioning approach is readily applicable.

References

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical-conditioning - Reply. *Behaviour Research and Therapy, 33*(7), 825-831.

Baeyens, F., Eelen, P., & Crombez, G. (1995). Pavlovian associations are forever - On classical-conditioning and extinction. *Journal of Psychophysiology, 9*(2), 127-141.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning - Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*(2), 133-142.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective evaluative value - Conservative but not unchangeable. *Behaviour Research and Therapy, 27*(3), 279-287.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1992). The content of learning in human evaluative conditioning - Acquired valence is sensitive to US-revaluation. *Learning and Motivation, 23*(2), 200-224.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*(5), 485-494.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*(2), 161-174.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative Learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853-869.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*(2), 245-268.

Dickinson, A., & Pearce, J. M. (1977). Inhibitory interactions between appetitive and aversive stimuli. *Psychological Bulletin, 84*(4), 690-711.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*(2), 345-355.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11,* 549-554.

Eertmans, A., Baeyens, F., & Van Den Bergh, O. (2001). Food likes and their relative importance in human eating behavior: Review and preliminary suggestions for health promotion. *Health Education Research, 16*(4), 443-456.

Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology - Animal Behavior Processes, 25*(2), 211-224.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology, 41*(6), 618-626.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the Implicit Association Test. *Journal of Consumer Research, 35*(1), 178-188.

Greenwald, A. G., & Nosek, B. A. (2001). Health of the Implicit Association Test at age 3. *Zeitschrift Fur Experimentelle Psychologie, 48*(2), 85-93.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34*(2-3), 115-139.

Hermans, D., De Houwer, J., & Eelen, P. (1996). Evaluative decision latencies mediated by induced affective states. *Behaviour Research and Therapy, 34*, 483-488.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion, 15*(2), 143-165.

Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion, 17*(1), 83-99.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (in press). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*.

Karpinski, A., & Hilton, J. L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology, 81*(5), 774-788.

Kerkhof, I., Goesaert, E., Dirikx, T., Vansteenwegen, D., Baeyens, F., D'Hooge, R., & Hermans, D. (2009). Assessing valence indirectly and online. *Cognition & Emotion, 23*, 1615-1629.

Kerkhof, I., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2009). A picture-flavor paradigm for studying complex conditioning processes in food preference learning. *Appetite, 53,* 303-308.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative learning in human pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239.

Mineka, S., & Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders - It's not what you thought it was. *American Psychologist, 61*(1), 10-26.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*(4), 421-433.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91*(6), 995-1008.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology, 37*, 867-878.

Sherer, K. R. (1993). Neuroscience projections to current debates in emotion psychology. *Cognition and Emotion, 7*, 1-42.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45.

Stevenson, R. J., Boakes, R. A., & Wilson, J. P. (2000). Counter-conditioning following human odor-taste and color-taste Learning. *Learning and Motivation, 31*(2), 114-127.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(1), 71-79.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition & Emotion*, *23*, 889-917.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognition & Emotion, 19*(2), 175-196.

experiment 4

The context-sensitivity of counterconditioning: An ABA renewal study

In the present experiment, we examined the context-sensitivity of evaluative and expectancy learning in a counterconditioning paradigm in which we compared a group of participants that received acquisition training, counterconditioning and test in a single context A, with a group that underwent acquisition training in context A, counterconditioning in context B and that subsequently returned to context A for test. For half of the participants, evaluative ratings were collected at the start, middle and end of each phase. The other participants were required to give US-expectancy ratings at these moments. At the end of the experiment, participants' evaluations were assessed indirectly in both contexts with an affective priming task. An ABA renewal effect was observed in the US-expectancy ratings but not in the evaluative ratings. Also in the priming data no evidence was obtained for contextual modulation of evaluative learning. Our findings thus suggest that, unlike expectancy learning, evaluative learning might not be sensitive to context manipulations.

Introduction

In the previous study, we found evaluative conditioning (EC) to be susceptible to counterconditioning (CC). By pairing the CS with a US that had a valence opposite from the original acquisition US, we succeeded in eliminating its previously acquired conditioned valence. This is an important finding, especially given that several studies (including our previous study) found EC to be little or not sensitive to extinction (e.g., Diaz, Ruiz, & Baeyens, 2005; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006).

In the general introductory chapter of this dissertation, we hypothesized that the extinction-resistance of evaluative learning might have important implications for the treatment of fear and phobias. These findings seem to suggest that a standard exposure intervention (the clinical analogue of extinction) might leave the acquired negative meaning of a phobic object relatively unaltered. As there is evidence that this remaining negative valence could form an affective-motivational source for the re-emergence of the original phobic fear (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Huijding & de Jong, 2005), we argued that it might be therapeutically beneficial to combine exposure with techniques aimed at altering the valence of the fear-eliciting stimulus. The results of the previous study suggest that CC provides a promising technique to alter an object's (conditioned) valence and therefore might function as an adjunct to mere exposure in the treatment of (return of) fear.

As mentioned earlier, hitherto, not much research has been conducted on evaluative CC and its properties. Procedurally, CC can be conceptualized as a retroactive interference paradigm in which participants learn conflicting associations in two temporally separated phases. Research on expectancy/signal learning showed that in interference paradigms (like CC but also extinction), the second learning phase does not involve the 'unlearning' or destruction of the first acquired association, but rather entails the learning of a new association which temporarily suppresses the former one. Evidence for this comes from studies on rapid reacquisition, spontaneous recovery, renewal, and reinstatement, phenomena that illustrate that the original CS-US association can quickly be restored with new learning trials or can suddenly re-emerge after the mere passage of time, a context change, or unsignaled presentations of the original US (for a review, see Bouton, 2002, 2004). Studies that focused on expectancy learning in interference paradigms further suggest that 'second-learned' information is more time- and context-dependent than first-learned information. According to Bouton (2002, 2004), our learning and memory system treats the first-learned association (e.g., CS-US or CS-US_1) as context-free, but the second association (e.g., CS-noUS or CS-US_2) as a kind of time- and context-specific 'exception to the rule'. An interesting question is whether second-learned *evaluations* (rather than *expectancies*) are also time- and context-specific. The answer to this question holds important implications for the use of CC techniques in clinical practice. If the new evaluations that patients acquire through CC are also found to be time- and context-specific, this would suggest that the possible beneficial therapeutic effects of a CC treatment might only be temporary. Participants' original negative evaluation of the fear object might reappear over time or upon leaving the therapeutic context.

In the previous study, we explored the *time-sensitivity* of CC by assessing valence at one week follow-up and found no evidence for spontaneous recovery.[21] In the present study, we explored the *context-sensitivity* of EC in an ABA renewal design in which we compared a group of participants that received preference acquisition, CC and test in a single context A (the 'AAA group'), with a group that received acquisition in context A, underwent CC in a different context (B), and then was tested for renewal in context A again (the 'ABA group'). We were interested in what would happen with participants' counterconditioned preferences in the ABA group upon return to the initial acquisition context. Will the effect of the CC treatment persist or will participants' initially (during the first acquisition phase) acquired evaluations reappear? Put differently: Will participants' evaluations become modulated by the contexts?

[21] Note that the terms 'spontaneous recovery' and 'renewal' are generally used to refer to the observation that an *extinguished* conditioned response can reappear over time or after a context switch (learning phase 1 = acquisition, learning phase 2 = extinction). In the present manuscript (and dissertation), we use these terms to refer to the possible reappearance of an initially acquired conditioned response after *counterconditioning* (learning phase 1 = acquisition, learning phase 2 = counterconditioning). In both cases, these terms refer to the reappearance of an initially learned conditioned response after a second learning phase in which a new response was acquired.

Only few studies have addressed the question of whether conditioned evaluations can become conditional upon the presence or absence of a certain cue stimulus (e.g., a context stimulus as suggested in the previous paragraph). Moreover, the studies that have examined whether 'modulation' or 'occasion setting'[22] can occur in EC have yielded conflicting results. In a series of experiments using the flavor-flavor paradigm, Baeyens and colleagues (Baeyens, Crombez, De Houwer, & Eelen, 1996; Baeyens, Hendrickx, Crombez, & Hermans, 1998) examined whether a discrete stimulus could become a modulator of participants' evaluatively conditioned flavor preferences. In their first experiment (Baeyens et al., 1996, Exp 1), the color of a drink (the 'feature') signaled whether a particular fruit flavor (the 'target') would be followed by a bad aftertaste. For instance, a fruit flavor was followed by the bad aftertaste only in green drinks but not in uncolored drinks. In this study, Baeyens et al. (1996) obtained no evidence for the acquisition of a color-modulated flavor-flavor association: participants acquired an unmodulated dislike for the fruit flavor. In several follow-up experiments (Baeyens et al., 1996, Exp 2-4; Baeyens et al., 1998), the authors explored whether modulated preferences could be obtained when a flavor (instead of color) cue was used as feature, when the number of acquisition trials was increased, when the feature and target were presented sequentially rather than simultaneously, and when the target was paired with a positive US (instead of no US) in the absence of the feature. In none of these experiments was evidence obtained for modulation of evaluative learning. Hardwick and Lipp (2000), however, did find occasion setting in a sequential feature positive EC procedure. In their procedure, participants were presented with a picture of a circle (or a tactile stimulus) as the target CS. This CS was followed by an aversive electrotactile stimulus if preceded by a tactile (or visual) feature stimulus, but not if presented alone. Blink startle during the target stimulus, elicited by an acoustic probe stimulus, was found to be larger after the presentation of the occasion setter than in its absence. According to Hardwick

[22] In this manuscript, the terms 'modulation' (Swartzentruber, 1995) and 'occasion setting' (Holland, 1992) are used interchangeably. With both terms we refer to the fact that some Pavlovian paradigms endow a stimulus (the 'modulator' or 'occasion setter') not (only) with a simple association with the US, but rather with the ability to modulate the functioning of other stimuli associated with that US (i.e., set the ocassion under which these other stimuli elicit conditional responses).

and Lipp (2000) these results indicate that the target was evaluated more negatively in the presence of the feature and therefore provide evidence for modulation of EC. An important drawback of this study is, however, that Hardwick and Lipp (2000) relied on blink startle modulation as the sole measure of affective learning. As we will discuss in more detail in Part 3 of this dissertation, several authors have argued that the startle response does not provide a good index of EC because it can also be affected by factors other than the valence of the CSs (e.g., arousal). Finally, one study in the domain of social psychology examined whether context-dependent attitudes could be created by presenting participants with evaluatively heterogeneous information in two different contexts (Rydell & Gawronski, 2009). In a series of experiments, Rydell and Gawronski (2009) asked participants to form an impression of a fictional target person named Bob based on written information about this person. The valence of this information (i.e., positive or negative) was dependent upon the color of the background stimulus (i.e., yellow or blue). The authors found participants' indirectly measured[23] attitudes to also vary as a function of the background color of the computer screen and hence, like Hardwick and Lipp (2000), obtained evidence in favor of modulation in evaluative learning.

In the present study, we explored the context-sensitivity of EC in the picture-taste paradigm that was also applied in the previous studies. The experiment included four experimental groups: an AAA-valence group, an ABA-valence group, an AAA-expectancy group and an ABA-expectancy group. Participants in the AAA groups received acquisition, CC and test for renewal in the same context A. Participants in the ABA groups, on the other hand, underwent acquisition in context A, CC in a different context B, and were tested for renewal in context A again. Different contexts were created by the display of images of different bakeries on the background screen (for a similar context manipulation, see Fonteyne, Vervliet, Baeyens, Hermans, & Vansteenwegen, 2009). For the valence groups, the main dependent variable were evaluative ratings while the participants of the expectancy groups provided US-expectancy ratings. Ratings were collected at the start, middle and end of each phase. The

[23] In all experiments, participants' evaluations were measured with the Affective Misattribution Procedure (AMP) of Payne, Cheng, Govorun, and Stewart, 2005. In one experiment (Exp 2), an affective priming task was included as well.

main aim of the present experiment was to explore the context-sensitivity of EC. The expectancy groups were included, however, as a manipulation check to verify that renewal effects could be obtained in our paradigm. As in previous studies on expectancy learning (e.g., Vansteenwegen et al., 2005; Rosas, Vila, Lugo, & Lopez, 2001), we expected to observe a renewal effect[24] for the ABA-expectancy group upon return to the original acquisition context.

Participants' evaluations were also assessed indirectly with an affective priming task (APT). After the conditioning procedure, all participants completed two APTs: One in the original acquisition context and one in the B context.

Method

Participants

Eighty-three naive psychology students (73 women) participated for partial fulfillment of course requirements or a monetary reward of € 9. Participants were randomly assigned to one of the four experimental groups: the AAA-valence group (n = 21), the ABA-valence group (n = 21), the AAA-expectancy group (n = 20) or the ABA-expectancy group (n = 21).

Materials

Two pictures of cookies, differing in color and shape, served as CSs. One cookie was a yellow triangle; the other was a red moon shape. They were centered on a grey background and had a displayed diameter of 5-6 cm. The total picture had a width of 512 pixels and a height of 384 pixels.

Good and bad tasting cookies served as USs. The pastry was made with 200 g flour, 100 g sugar ('good' cookies only), one egg and 125 g butter. It was divided in two halves and by adding two tablespoons of hazelnut or four tablespoons of Tween20 (*polysorbate 20*; a bitter soapy tasting chemical) 'intended' good, respectively bad, tasting cookies were obtained. All cookies were yellow squares with a side length of 1 cm. The US-cookies were wrapped in aluminum foil to make sure that participants would not be able to distinguish the good and bad tasting cookies by sight. CS-US pairings were counterbalanced over participants.

[24] I.e., a return of participants' acquisition expectancies.

Two similar but easily distinguishable pictures of bakeries were used as experimental contexts (see Figure 5). The CS-pictures were presented in the middle of the computer screen superimposed on one of these two context slides. Context pictures were 1024 x 768 pixels (i.e., screen-filling). Which picture served as context/bakery A versus context/bakery B was counterbalanced over participants.

Figure 5. Experimental material used in Experiment 4: context pictures.

For the APT, the CS-pictures served as primes, and ten positive and ten negative Dutch nouns (selected from Hermans & De Houwer, 1994) were used as targets.

An Affect 4.0 program (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010) controlled stimulus presentation and response registration.

Procedure

Participants were tested in a lab containing ten identical separated test boots, arranged in a semicircle, such that visual contact between participants was precluded, but each participant could clearly see and hear the experimenter. The experimenter was seated behind a table at the centre of the semicircle.

Upon entering the lab, participants received a short introduction about the experiment and were asked to fill out an informed consent form. Subsequently, they were asked to take place in one of the test boots. In this test room, participants were presented with a cup of water and a piece of plastified paper on which 44 small cookies (wrapped in aluminum foil) were arranged in two rows of 22.[25] At the beginning of each row was a small picture of one of the CSs. After participants had taken a seat, they were instructed about the upcoming conditioning procedure, which consisted of three phases: an acquisition phase, a CC phase and a test phase.

The *acquisition phase* consisted of two blocks of 6 trials. Within each block, every CS was presented three times. A conditioning trial proceeded as follows (also see Table 3): participants were first presented with one of the two CS-pictures on the computer screen. One thousand ms after CS onset, the experimenter instructed them to press the space bar, whereupon the message '*Take and eat a cookie*' appeared on the screen. At that moment participants were required to eat a cookie from the row containing the presented CS-picture. Twenty seconds later, participants were asked to take a sip of water to avoid aftertastes. After another 20 s, the experimenter instructed participants to press the enter key to proceed to the next trial. During the acquisition phase, one CS was consistently paired with positive US cookies, while the other one was consistently paired with bad tasting cookies. Presentation order was randomized for every participant, with the restriction that no CS-picture appeared more than twice in a row. The intertrial interval (ITI) was 5 s.

The context in the conditioning procedure was manipulated by the background of the computer screen. During the entire acquisition phase (trials + ITI), the picture of bakery A was used as background screen.

Self-report ratings were collected at the start, middle (i.e., after the first acquisition block) and end of the acquisition phase. Half of the participants (the valence groups, n = 42), were asked to rate the (un)pleasantness of the CSs on a scale ranging from -100 (*very unpleasant*) to +100 (*very pleasant*). This scale was accompanied by the question: 'How pleasant/unpleasant do you evaluate this picture *at this moment*?'. The other half of the participants (the expectancy groups, n = 41) were required to rate their expectancy of the positive or negative US after each CS. The expectancy rating scale was accompanied by the question: 'To what extent do you expect this picture to be followed by a hazelnut

[25] Even though there were only 36 conditioning trials, we presented the participants with 44 cookies to prevent them from being able to deduce from the number of cookies when the conditioning procedure would end (as this might, for instance, influence their expectancies).

or tween cookie *at this moment*?' and ranged from -100 (*I certainly expect a tween cookie)* to +100 (*I certainly expect a hazelnut cookie*). A rating trial proceeded as follows (also see Table 4): like in an acquisition trial, participants were presented with a CS-picture that appeared against the background of bakery A. Thousand ms after CS onset, participants were requested to fill out a rating scale. Ten seconds later, the experimenter instructed participants to press the space bar, whereupon the message 'This time you do not have to eat a cookie' appeared on the screen. Subsequently, participants were asked to press the enter key to proceed to the next (rating or conditioning) trial. Participants always completed two ratings: one for each CS. The intertrial interval (ITI) was 5 s.

Table 3
Sequence and timing of events within a conditioning trial

Time within trial	**Event(s)**
0 ms	Onset CS
1000 ms	Participants are instructed to press the space bar Onset message 'Take and eat a cookie' Participants take and eat a cookie
21000 ms	Participants are instructed to take a sip of water
41000 ms	Participants are instructed to press ENTER to proceed to the next trial (Offset CS, start ITI)
46000 ms	End ITI

Note. Depending on the conditioning phase, bakery A or B was presented as a background screen during the conditioning trials and ITI.

Table 4
Sequence and timing of events within a rating trial

Time within trial	**Event(s)**
0 ms	Onset CS
1000 ms	Participants are instructed to fill out a rating scale
11000 ms	Onset message 'This time you do not have to eat a cookie' Participants are instructed to press ENTER to proceed to the next trial (start ITI)
16000 ms	End ITI

Note. Depending on the conditioning phase, bakery A or B was presented as a background screen during the rating trials and ITI.

The *CC phase* immediately followed the acquisition phase and consisted of 2 blocks of 12 trials. Within each block, every CS was presented six times. In this phase, the CS-US contingencies were reversed. Hence, the CS that was paired with good tasting cookies during acquisition was now paired with bad tasting cookies and vice versa. For half of the participants (AAA-valence group and AAA-expectancy group), the same context picture was used as during acquisition. For the other half (ABA groups), bakery B functioned as background picture during the CC phase. The course of the trials and presentation parameters were identical as for the acquisition phase.

Valence (valence groups) or US-expectancy ratings (expectancy groups) were also collected at the start, middle and end of this phase. For the ABA (but not AAA) groups, bakery B was used as background picture during these rating trials.

The *test phase* consisted of two rating trials; one for each CS. For all participants, the acquisition bakery (i.e., bakery A) functioned as background picture during this phase. Hence, for the ABA (but not AAA) groups, the test phase entailed a context switch.

Table 5 provides an overview of the experimental procedure for each group. Because ratings were collected at the beginning and end of each phase, there were two moments in the experiment at which participants were required to fill out four instead of two rating scales: in between acquisition and CC and at the end of the experiment. To prevent participants (and in particular the AAA participants[26]) from interpreting these double rating blocks as a cue for contingency reversal, we decided to collect four ratings (two per CS) at each test moment (see Table 5).

[26] As there is no context cue that makes the contingency reversal understandable, these participants may be inclined to look for other predictive cues.

Table 5
Design of Experiment 4

	Rating	Acquisition first half	Rating	Acquisition second half	Rating	Rating	CC first half	Rating	CC second half	Rating	Test
AAA valence	val (4)	CS_1 - L (3) CS_2 - D (3)	val (4)	CS_1 - L (3) CS_2 - D (3)	val (2)	val (2)	CS_1 - D (6) CS_2 - L (6)	val (4)	CS_1 - D (6) CS_2 - L (6)	val (2)	val (2)
ABA valence	val (4)	CS_1 - L (3) CS_2 - D (3)	val (4)	CS_1 - L (3) CS_2 - D (3)	val (2)	val (2)	CS_1 - D (6) CS_2 - L (6)	val (4)	CS_1 - D (6) CS_2 - L (6)	val (2)	val (2)
AAA expectancy	exp (4)	CS_1 - L (3) CS_2 - D (3)	exp (4)	CS_1 - L (3) CS_2 - D (3)	exp (2)	exp (2)	CS_1 - D (6) CS_2 - L (6)	exp (4)	CS_1 - D (6) CS_2 - L (6)	exp (2)	exp (2)
ABA expectancy	exp (4)	CS_1 - L (3) CS_2 - D (3)	exp (4)	CS_1 - L (3) CS_2 - D (3)	exp (2)	exp (2)	CS_1 - D (6) CS_2 - L (6)	exp (4)	CS_1 - D (6) CS_2 - L (6)	exp (2)	exp (2)

Note. Val = valence, exp = expectancy, L = liked US, D = disliked US. Numbers in brackets indicate number of trials. White areas indicate learning trials in context A, grey areas indicate learning trials in context B.

After the conditioning phases, participants completed two APTs: one in the original acquisition context (i.e., bakery A was used as background picture during the entire task) and one in the B context (i.e., bakery B was used as background picture). The experimenter explained that on each trial a picture (called the 'prime') would precede a word (called the 'target'). Participants were instructed to attend the word and evaluate it as quickly as possible by pressing the right, respectively left, mouse button for positive, respectively negative, words. The priming task consisted of 40 experimental trials, preceded by 12 practice trials. Each trial started with a fixation cross (500 ms), followed by a 500 ms time gap, followed by the prime (200 ms). Fifty ms after prime offset, the target was presented until participants responded or 2000 ms elapsed. All stimuli were presented against the background of bakery A (first APT) or B (second APT). The mean ITI was 1000 ms (range 500-1500 ms). During the ITI, the bakery pictures remained on the screen. Every prime was paired with ten negative and ten positive targets. Trial order was random with the restriction that the same prime could never be presented on more than three subsequent trials.

After the APTs, participants from the ABA (but not AAA) groups were instructed to write down any differences they had noticed between the first and second part of the conditioning procedure. Subsequently, they were asked whether they had noticed that the background of the computer screen was different for these two phases (*yes/no*). The ABA participants were then presented with both bakery pictures and were asked to indicate for each conditioning phase which picture served as background stimulus and how certain they were of this (*very uncertain, reasonably uncertain, reasonably certain, very certain).*

Contingency recall was assessed for all participants by asking them to indicate for each CS for both conditioning phases whether it had been followed by a hazelnut or Tween20 cookie and how certain they were of this *(very uncertain, reasonably uncertain, reasonably certain, very certain)*. Finally, all participants rated the US flavors on a scale ranging from -100 *(very unpleasant)* to +100 *(very pleasant)*.

Results

US ratings

A prerequisite for EC to occur is that the used USs are experienced as intended by the participants. The analysis of the US ratings revealed that this precondition was met. As expected, the intended positive US was rated significantly more positive (M = 72.05, SD = 26.58) than the intended negative US (M = -61.33, SD = 30.23), $t(82)$ = 28.67, p < .0001.

CS ratings

Data analysis

Participants' self-report ratings were analyzed using a 2 (Group: AAA vs. ABA) x 2 (CS-type: LD vs. DL[27]) x 7 (Moment: start, middle, end of acquisition phase; start, middle, end of CC phase[28]; test phase) ANOVA with group as between-subjects factor and CS-type and moment as within-subjects factors.

To assess the presence of acquisition, CC and renewal effects, we performed a series of planned comparisons. The reliability of *acquisition* was evaluated by testing a CS-type x Moment (start vs. end of acquisition) interaction across groups (to establish acquisition overall) and by a Group x CS-type x Moment interaction (to check whether the acquisition effect differed between the AAA and ABA groups). If the latter interaction was significant, we looked at the reliability of acquisition within each group separately. To explore whether acquisition *generalized* to the next phase, we compared participants' ratings at the end of acquisition with their ratings at the start of the CC phase by means of a CS-type x Moment (end of acquisition vs. start of CC) interaction across groups. If the Group x CS-type x Moment interaction (to test whether the generalization differed between the two groups) was significant, we also looked

[27] LD = CS that was paired with a liked US during the acquisition phase and a disliked US during the counterconditioning phase; DL = CS that was paired with a disliked US during acquisition and a liked US during the counterconditioning phase.

[28] As explained above, double rating blocks were inserted at each test moment to prevent participants from interpreting the double ratings as a cue for contingency reversal. As a result of this, we have two ratings for each CS (LD/DL) at the start of the acquisition phase, after the first acquisition block and after the first counterconditioning block (also see Table 5). The presented analyses were conducted on the average of these both ratings. A similar pattern of results was obtained, however, when participants' first or second rating scores were used for these test moments.

at generalization of acquisition within each group separately. To evaluate *CC*, we calculated a CS-type x Moment (start vs. end of CC) interaction across groups. To check whether the effect of CC differed between both groups, we also calculated a Group x CS-type x Moment interaction. If the latter interaction was significant, we examined the CS-type x Moment (start vs. end of CC) interaction within each group separately. To test for *renewal*, we compared participants' ratings at the end of the CC phase with their ratings in the test phase by means of a Group x CS-type x Moment (end of CC vs. test phase) interaction. If this interaction was significant, indicating a difference between the groups in renewal of conditioned responding, we followed up on this analysis with simple CS-type x Moment (end of CC vs. test phase) interactions within each group.

Expectancy ratings

One participant from the AAA-expectancy group failed to comply with the rating instructions and therefore was excluded from the subsequent analyses. Participants' mean US-expectancy ratings as a function of CS-type and moment are depicted in Figure 6.

<u>Acquisition.</u> We expected participants to gradually acquire the CS-US contingencies throughout the acquisition phase. We predicted to observe little or no differentiation in participants' expectancy ratings for both CSs (LD/DL) at the beginning of acquisition, but expected to observe a positive rating (indicating the expectancy of a hazelnut cookie) for the LD stimulus and a negative rating (indicating the expectancy of a Tween20 cookie) for the DL stimulus at the end of the acquisition phase. Since all participants (AAA and ABA) underwent the same acquisition treatment, we predicted a similar pattern of results in both groups.

Planned comparisons revealed a significant CS-type x Moment (start vs. end acquisition) interaction, $F(1, 38) = 224.14$, $p < .0001$. Across both groups, participants learned to expect the positive US after the LD stimulus and the negative US after the DL stimulus (also see Figure 6). Unexpectedly, this differential acquisition effect was more pronounced in the ABA-expectancy group, as reflected by a significant Group x CS-type x Moment (start vs. end acquisition) interaction, $F(1, 38) = 8.08$, $p < .01$. Importantly, the CS-type x Moment interaction was significant in both experimental groups, with $F(1, 38) = 70.05$, $p < .0001$ for the AAA group and $F(1, 38) = 167.01$, $p < .0001$ for the ABA group. Hence, successful differential acquisition was obtained in both groups.

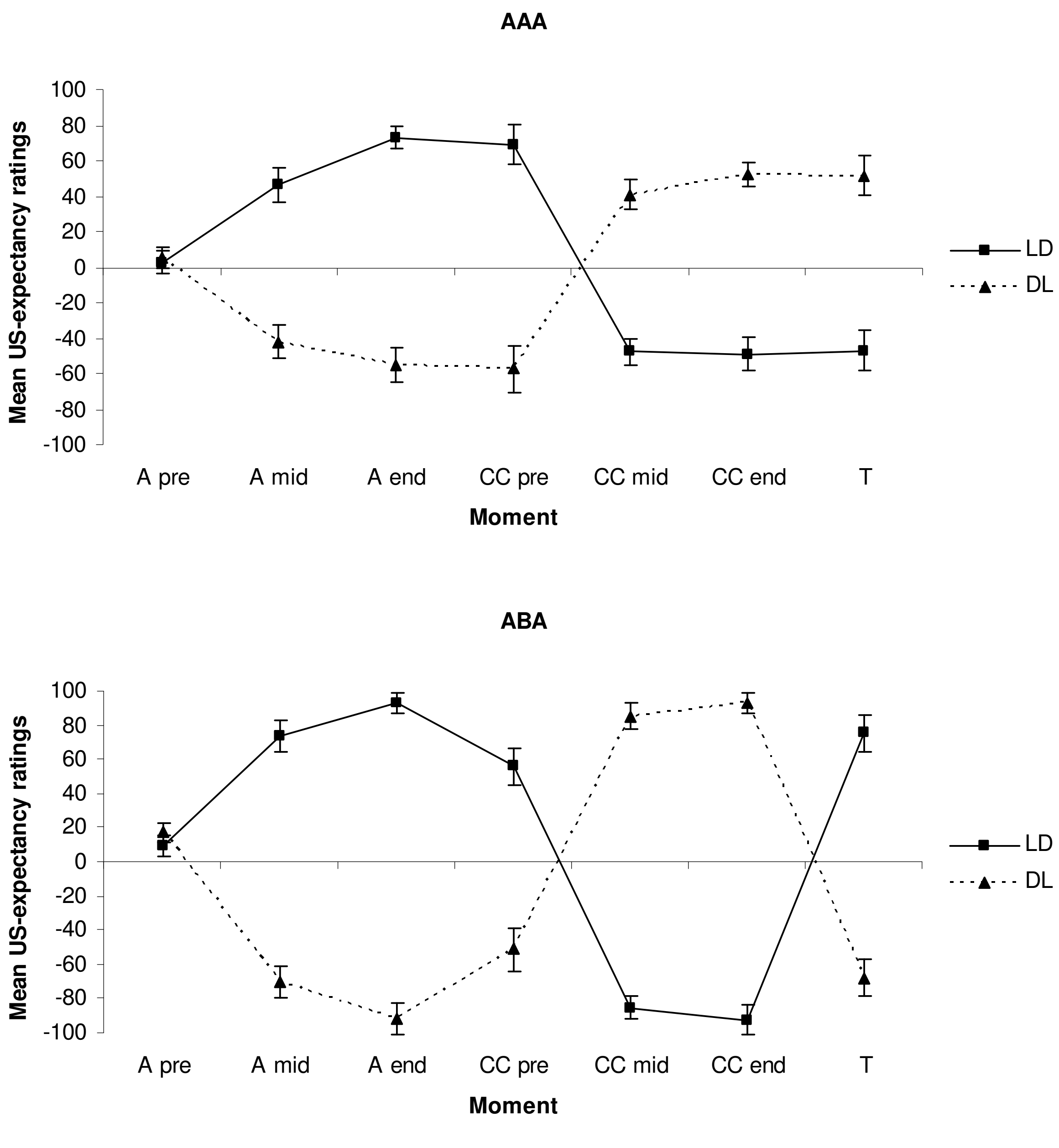


Figure 6. Mean US-expectancy ratings (+SE) for the AAA (upper panel) and ABA (lower panel) groups by CS-type and moment. The US-expectancy scale ranged from -100 (*I certainly expect a tween cookie*) to +100 (*I certainly expect a hazelnut cookie*).

Generalization of acquisition. We expected participants to generalize what they had learned during acquisition to the CC context.

Planned comparisons, however, revealed a significant CS-type by moment (end acquisition vs. start CC) interaction, $F(1, 38) = 10.69$, $p < .01$. Across groups, the acquired differentiation in US-expectancy was found to decrease upon entering the CC context. The degree of generalization seemed to differ, however, between both experimental groups, as suggested by a significant Group x CS-type x Moment (end acquisition vs. start CC) interaction, $F(1, 38) = 9.59$, $p < .01$. Indeed, simple CS-type x Moment interactions within each group revealed a preserved differential expectancy effect for the AAA group, $F < 1$, but some loss in differentiation for the ABA group, $F(1, 38) = 21.34$, $p < .0001$. Importantly, a pairwise comparison between CS-types (LD vs. DL) at the start of the CC phase in the ABA group still revealed a significant difference in participants' US-expectancy ratings for these two stimuli (with a positive rating for the LD stimulus and a negative rating for the DL stimulus), $F(1, 38) = 24.38$, $p < .0001$. Hence, even though the context switch resulted in some generalization decrement, participants of the ABA group still expected the acquisition contingencies to hold in the CC context.

Counterconditioning. We expected participants to gradually acquire the new contingencies during the CC phase and predicted them to expect the hazelnut cookie after the DL stimulus and the Tween20 cookie after the LD stimulus at the end of this phase.

Planned comparisons indeed revealed a significant CS-type by moment (start vs. end CC) interaction, $F(1, 38) = 167.27$, $p < .0001$. Whereas at the start of the CC phase, participants still expected the acquisition contingencies to hold, they indicated expecting the positive US after the DL stimulus and the negative US after the LD stimulus at the end of the CC phase. The nonsignificant Group x CS-type x Moment (start vs. end CC) interaction, $F(1, 38) = 2.62$, *ns*, indicated that the groups did not differ in degree of CC.

Renewal of conditioned responding. Based on previous studies that found expectancy learning to be sensitive to ABA renewal, we expected to observe a renewal effect (i.e., positive ratings for the LD stimulus and negative ratings for the DL stimulus) in the ABA group (for which the renewal phase entailed a context switch) but not in the AAA group (for which there was no context switch).

Confirming this prediction, planned comparisons demonstrated a significant Group x CS-type x Moment (end CC vs. test phase) interaction, $F(1, 38) = 160.80$, $p < .0001$. Simple CS-type x Moment (end CC vs. test phase) interactions conducted within each group revealed a renewal effect in the ABA group, $F(1, 38) = 344.02$, $p < .0001$, but not in the AAA group, $F < 1$. As expected, participants in the AAA group still expected the contingencies of the CC phase to hold and demonstrated positive ratings (indicating the expectancy of a hazelnut cookie) for the DL stimulus and negative ratings (indicating the expectancy of a Tween20 cookie) for the LD stimulus. The participants of the ABA group, however, expected the acquisition contingencies to hold again and showed a reversed pattern of results (i.e., positive ratings for the LD stimulus and negative ratings for the DL stimulus), indicative of renewal.

In sum, we obtained successful differential acquisition, generalization and CC in the US-expectancy ratings of the participants of the AAA and ABA-expectancy groups. Expectancy learning was further found to be context sensitive as a renewal effect was observed for the ABA but not AAA group upon return to the original acquisition context. Based on these data, we can also conclude that our use of different background pictures worked as a context manipulation.

Valence ratings

Participants' mean valence ratings as a function of CS-type and moment are depicted in Figure 7.

<u>*Acquisition.*</u> We expected the CS that was paired with positive cookies during the acquisition phase (i.e., the LD stimulus) to acquire a positive meaning and the CS that was combined with negative cookies (i.e., the DL stimulus) to become negative. As all participants (AAA and ABA) underwent the same acquisition procedure, we expected a similar pattern of results in both experimental groups.

Planned comparisons revealed a significant CS-type by moment (start vs. end acquisition) interaction, $F(1, 40) = 75.17$, $p < .0001$. As predicted, the LD stimulus acquired a positive connotation because of its pairing with the positive US. The DL stimulus, on the other hand, became negative during the acquisition phase (also see Figure 7). This EC effect was equally strong in both groups, as reflected by a nonsignificant Group x CS-type x Moment (start vs. end acquisition) interaction, $F < 1$.

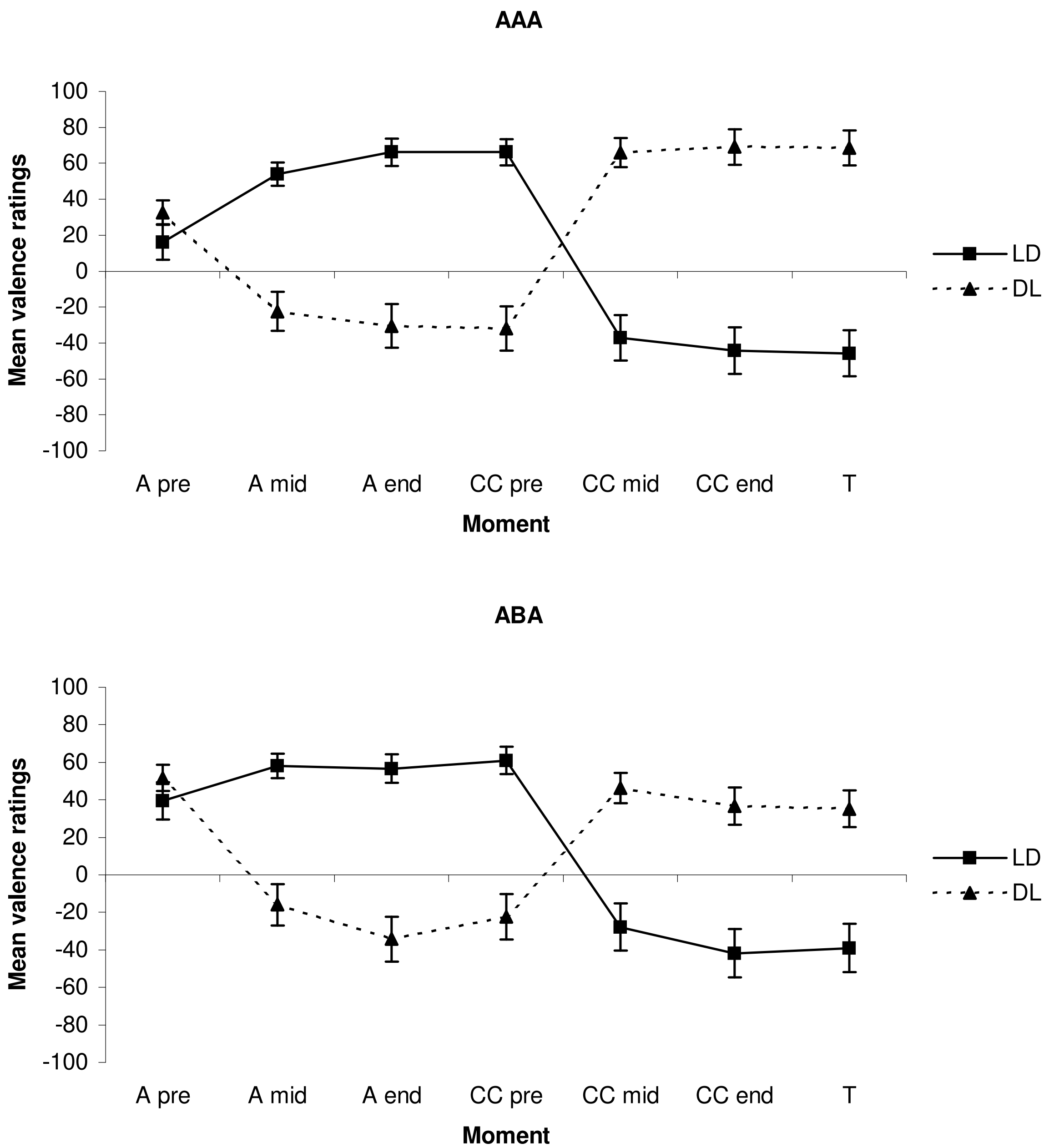


Figure 7. Mean valence ratings (+SE) for the AAA (upper panel) and ABA (lower panel) groups by CS-type and moment. The evaluative rating scale ranged from -100 (*very unpleasant)* to +100 (*very pleasant*).

Generalization of acquisition. Also for the valence ratings, we expected participants to generalize what they had learned during acquisition to the CC context.

In line with this prediction, planned comparisons demonstrated a nonsignificant CS-type by moment (end acquisition vs. start CC) interaction, $F(1, 40) = 1.86$, *ns*. The Group x CS-type x Moment (end acquisition vs. start CC) interaction was, however, marginally significant, $F(1, 40) = 3.98$, $p = .05$, suggesting some difference in the degree of generalization between both groups. Simple CS-type x Moment interactions within each group revealed no generalization decrement in the AAA group, $F < 1$, but a small decrease in the EC effect of the ABA group from the end of the acquisition phase to the start of the CC phase, $F(1, 40) = 5.64$, $p < .05$. Importantly, a pairwise comparison between CS-types (LD vs. DL) at the start of the CC phase in the ABA group still revealed a significant EC effect (with a positive rating for the LD stimulus and a negative rating for the DL stimulus), $F(1, 40) = 32.27$, $p < .0001$.

Counterconditioning. We expected participants' evaluations to gradually alter over the CC phase. At the end of this phase, we expected participants to evaluate the LD stimulus negatively and the DL stimulus positively (i.e., in line with the contingencies of the CC phase).

The CS-type by moment (start vs. end CC) interaction was highly significant, $F(1, 40) = 96.54$, $p < .0001$. As predicted, the LD stimulus went from a positive to a negative connotation during the CC phase, while the reverse pattern was observed for the DL stimulus. The nonsignificant Group x CS-type x Moment (start vs. end CC) interaction, $F(1, 40) = 1.70$, *ns*, indicated that the groups did not differ in degree of CC.

Renewal of conditioned responding. To test for a possible renewal effect in the evaluative ratings, we conducted a Group x CS-type x Moment (end CC vs. test phase) interaction. This interaction, however, failed to approach significance, $F(1, 40) = 2.35$, *ns.* As can be seen in Figure 7, participants' evaluations of the CSs remained the same at both test moments, both in the AAA and ABA groups.

To conclude, we obtained successful differential acquisition, generalization and CC in the valence ratings of the participants of the AAA and

ABA-valence groups. Unlike in the US-expectancy ratings, however, no renewal effect was observed in these data. Participants' self-reported ratings thus suggest that, unlike expectancy learning, evaluative learning is not sensitive to context modulation.[29]

Priming data

Data analysis

Participants who made 20% or more errors on the APT were excluded from the analysis.[30] This resulted in the exclusion of 11 participants (one of the AAA-valence group, three of the ABA-valence group, six of the AAA-expectancy group and one of the ABA-expectancy group). The data were corrected for outliers by excluding reaction times (RTs) below 200 ms or above 1500 ms (1.42%). Trials with no (0.01%) or incorrect responses (8.37%) were also discarded. To create an evaluation score, the mean latency for positive target words was subtracted from the mean latency for negative words for each prime category (Gawronski, Walther, & Blank, 2005). Higher scores on this variable indicate more positive evaluations. Note that these scores should not be interpreted in an absolute manner (e.g., a value of zero reflecting a neutral evaluation), because response latencies for positive target words may generally differ from response latencies for negative target words.

A preliminary analysis revealed no main or interaction effects involving the variable measure (US-expectancy ratings vs. valence ratings). Therefore, the priming data were pooled across this variable. Participants' priming scores were analyzed using a 2 (Context: A vs. B) x 2 (Group: AAA vs. ABA) x 2 (CS-type: LD vs. DL) repeated measures ANOVA with group as between-subjects factor and CS-type and context as within-subjects factors.

Before presenting the results of this analysis, we briefly discuss our predictions for the priming data. If evaluative learning would be *insensitive to*

[29] This conclusion was confirmed by a 2 x 2 x 2 x 2 MANOVA with measure (US-expectancy ratings vs. valence ratings) and group (AAA/ABA) as between-subjects factors and CS-type (LD/DL) and moment (end of counterconditioning vs. renewal phase) as within-subjects factors. This analysis revealed a significant Measure x Group x CS-type x Moment interaction, $F(1, 78) = 160.73$, $p < .0001$, indicating that the renewal effect differed for the US-expectancy and valence groups.

[30] Inclusion of these participants did not affect the overall pattern of results.

context modulation, we would expect to observe similar effects for the AAA and ABA groups in both APTs. In both groups, and both contexts (bakery A/B), we would expect to observe evaluations in line with the contingencies of the CC phase (as this phase was twice as long as the acquisition phase). Hence, in both APTs and for both groups, we would predict to observe a more positive evaluation score for the DL stimulus than for the LD stimulus. If, however, evaluative learning would be *susceptible to context modulation*, a different pattern of results can be expected for the AAA and ABA group upon return to the A context. A renewal effect would translate itself into more a more positive evaluation of the LD stimulus as compared to the DL stimulus in the ABA group. In the AAA group, no such renewal effect is expected as for this group both conditioning phases (acquisition and CC) took place in the same context (bakery A). Hence, in this group we expected to see more a more positive score for the DL stimulus than for the LD stimulus. In the B context, a similar pattern of results is expected for both experimental groups. In this context, we expect to observe evaluations in line with the contingencies of the CC phase (i.e., more positive scores for DL than LD) in both the AAA and ABA groups.

Priming results

Participants' mean APT scores for the LD and DL stimulus as a function of group and context can be seen in Figure 8.

The Context x Group x CS-type interaction failed to approach significance, $F < 1$, suggesting a similar pattern of results in both APTs (and thus both contexts). Even though this interaction was not significant, we nevertheless explored the pattern of results in each APT separately.

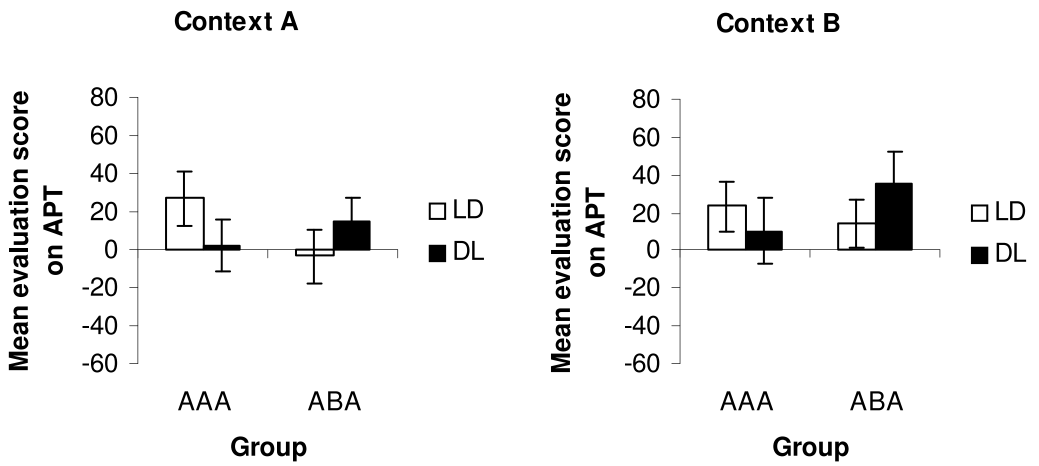


Figure 8. Mean evaluation score on the APT for the LD and DL stimulus as a function of group and context (left panel: context A, right panel: context B).

APT in Context A (Figure 8, left panel). In the first APT task, the Group x CS-type interaction was not significant, $F(1, 72) = 2.57$, *ns,* suggesting a similar pattern of priming results for both experimental groups in the A context. Unexpectedly, there was no difference in participants' priming scores for the LD and DL stimulus, both in the AAA and ABA groups, with $F(1, 72) = 1.71$, *ns* and $F < 1$.

APT in Context B (Figure 8, right panel). The Group x CS-type interaction also failed to approach significance in the second APT, $F(1, 72) = 1.28$, *ns*. Hence, similar priming results were also obtained for both experimental groups in the B context. Also in this task, participants' evaluation of the LD stimulus did not differ from their evaluation of the DL stimulus, both in the AAA and ABA groups, with both F's$(1, 72) = 1.02$, *ns.*

In sum, the priming analysis revealed similar results for both experimental groups in both APTs. Hence, like in the evaluative rating data, we found no evidence for contextual modulation of evaluative learning. Unexpectedly, however, in both groups and both APTs we found no difference in participants' priming scores for the LD and DL stimulus, which suggests that the value that these CSs acquired during the acquisition phase was abolished but not reversed by the CC treatment. This finding is surprising, given that the CC phase was twice as long as the acquisition phase, and stands in sharp contrast with the evaluative rating data, which demonstrated a fast and strong reversal of participants' CS evaluations during the CC phase.

Post-conditioning questionnaire

Context manipulation

Six of the 21 participants in the ABA-expectancy group and eight of the 21 participants in the ABA-valence group reported having noticed the context change in the open response questionnaire. A Chi-square test indicated no difference in the proportion of participants that spontaneously brought up the context switch between both groups, $\chi^2(1) = .43$, *ns.*

In the closed questionnaire, all but one participant (from the ABA-expectancy group) indicated having noticed the different background screens. Twenty of the 21 participants of the ABA-expectancy group and 18 of the 21 participants of the ABA-valence group correctly identified[31] which bakery was presented during the first and which during the second conditioning phase. A Chi-square test indicated no difference in the proportion of participants that correctly identified the two conditioning contexts between both groups, $\chi^2(1) = 1.11$, *ns.* Exclusion of the participants who did not notice the context switch or did not correctly identify which context was applied for which phase, did not affect the results.

Contingency recall

A participant was considered to remember a certain CS-US association when he/she correctly identified the US that had been paired with a specific CS, and was at least 'reasonably certain' of his/her choice. The mean numbers of recalled pairings for each experimental group are presented in Table 6.

[31] Participants had to choose the correct bakery picture and had to be at least reasonably certain of their choice.

Table 6
Mean number of recalled CS-US pairings (maximum = 2) and standard deviations as a function of group and moment

	Acquisition phase		***Counterconditioning phase***	
	M	***SD***	***M***	***SD***
AAA-valence group	1.05	0.97	1.62	0.74
ABA-valence group	1.62	0.80	1.71	0.72
AAA-expectancy group	.90	1.02	1.45	.89
ABA-expectancy group	2	0.00	2	0.00

A 2 (Group: ABA, AAA) x 2 (Measure: valence, expectancy) x 2 (Contingencies: acquisition, CC) ANOVA revealed main effects of group and contingencies, with $F(1, 79) = 19.65$, $p < .0001$ and $F(1, 79) = 9.49$, $p < .01$. In general, participants of the ABA groups remembered more contingencies than participants of the AAA groups. Also, the contingencies of the CC phase were remembered better than the contingencies of the acquisition phase. In addition to these main effects, there was a significant Group x Contingencies interaction, $F(1, 79) = 6.75$, $p < .05$. Whereas participants of the ABA groups remembered the acquisition and CC contingencies equally well ($M_{acq} = 1.81$, $SD_{acq} = 0.59$; $M_{cc} = 1.86$, $SD_{cc} = 0.52$), $F < 1$, participants of the AAA groups remembered fewer contingencies of the first conditioning phase than of the second phase ($M_{acq} = 0.98$, $SD_{acq} = 0.99$; $M_{cc} = 1.54$, $SD_{cc} = 0.81$), $F(1, 79) = 15.93$, $p < .0001$. All other main and interaction effects failed to approach significance.

Discussion

In the present experiment, we examined the context-sensitivity of evaluative and expectancy learning in a CC paradigm in which we compared a group of participants that received acquisition training, CC training and test in a single context A, with a group that underwent acquisition training in context A, CC in context B and that subsequently returned to context A for test.

The US-expectancy ratings revealed the expected pattern of results: strong differential acquisition which generalized well to the start of the CC phase and which was reversed by the end of the CC phase. Upon return to the original acquisition context, renewed responding was observed in the ABA but not AAA group. This finding is in line with several previous studies that found expectancy learning to be sensitive to context modulation in an ABA renewal design (e.g., Vansteenwegen et al., 2005; Rosas et al., 2001). The results of the expectancy ratings thus reassure us that renewal effects *can* be obtained in the paradigm we applied in the present experiment.

The evaluative ratings also revealed successful differential acquisition, generalization and CC. Unlike in the US-expectancy ratings, however, no renewal effect was observed in the ABA group. These data thus suggest that, unlike expectancy learning, evaluative learning is not sensitive to context modulation. In the ABA-group, a small but significant generalization decrement was observed from the end of the acquisition phase to the start of the CC phase. This might be seen as evidence that EC to some extent is sensitive to context manipulations. Note, however, that this effect disappeared when we excluded three participants in the ABA-valence group that showed no acquisition effect (i.e., an increase in the differentiation between the LD and DL stimulus from the start to the end of acquisition with the LD stimulus becoming more positive and the DL stimulus more negative) in their evaluative ratings.[32] When these participants were excluded, both the Group x CS-type x Moment (end acquisition vs. start CC) interaction and the simple CS-type x Moment interaction in the ABA group became nonsignificant, both *F's* < 1. These additional analyses thus suggest that the earlier reported generalization decrement resulted from a small minority of the participants for who EC failed all together. Participants who showed EC effects, revealed a perfect transfer of their acquired preferences to the B context.

[32] Five participants (3 in the ABA-valence group and 2 in the AAA-expectancy group) failed to show an acquisition effect in their ratings. Apart from the here reported difference, all other results remained the same when these participants were excluded from the rating or priming analyses. Three participants (2 in the ABA-valence group and 1 in the AAA-expectancy group) failed to show a counterconditioning effect in their ratings (i.e., a decrease or reversal in the differentiation between the LD and DL stimulus from the start to the end of the counterconditioning phase). Exclusion of these participants did also not change the reported results. Of the three participants that failed to show an effect of the counterconditioning treatment, two also failed to show an acquisition effect.

Also in the priming data no evidence was obtained for contextual modulation of evaluative learning. A similar pattern of results was obtained for the AAA and ABA groups, both in the A and B contexts. Unexpectedly, however, we found no difference in participants' priming scores for the LD and DL stimulus, which suggests that the differential value that these CSs were assumed to acquire during the acquisition phase was abolished but not reversed by the CC treatment. This finding is surprising, given that the CC phase was twice as long as the acquisition phase, and stands in sharp contrast with the evaluative rating data, which demonstrated a fast and strong reversal of participants' CS evaluations during the CC phase. As the priming data only revealed null effects, these results should be interpreted with caution. Because the APT was only administered at the end of the experiment, and not in between the acquisition and CC phase, we cannot exclude the possibility that the found results might indicate that our priming task simply failed to pick up participants' evaluations. Perhaps the bakery background interfered with the priming task, for instance by distracting participants' attention from the prime stimuli or categorization task. The high percentage (13%) of participants with 20% or more errors on the APT might also be understood in this light. In the only other study that manipulated the context in a priming task through the background screen (Rydell & Gawronski, 2009), the applied contexts were much simpler (i.e., a plain blue or yellow background screen) and hence less distracting. Moreover, the context screens were not presented during the entire priming task, but only appeared together with the prime stimuli. More research is needed to accurately interpret the priming data. It would, for instance, be interesting to replicate the present experiment but to administer an APT immediately after the acquisition phase to see if EC effects can be obtained in APT that takes place against the background of a complex picture. An alternative would be to use more simple backgrounds like in the Rydell and Gawronski (2009) study or to present the context pictures only during the priming presentations. The latter solution may, however, also not work for complex background pictures as participants might not be able to discriminate both context pictures during the short (200 ms) prime presentations.

An important disadvantage of the fact that our priming data are difficult to interpret, is that our conclusion that EC might be insensitive to context manipulations is mainly based on self-report data. This leaves open the possibility that our results could be driven by demand effects. It is not clear, however, why participants would believe the experimenter to hold different predictions for evaluative learning and expectancy learning. As Baeyens et al. (1996) already pointed out, in daily life, many instances can be found of preferences that seem to be conditional rather than absolute or free of contextual control. For example, a glass of brandy may be a delight after a luxurious dinner, but evoke a strong dislike immediately after breakfast cereals. Hence, it seems equally logic that participants would expect the experimenter to predict context-specific preferences. Although it is clear that there are problems associated with (retrospective) verbal reports, it might nevertheless be informative for future research to assess participants' knowledge of the experimental hypotheses and to ask them about the reasons behind their ratings (see also De Houwer, Baeyens, & Field, 2005).

The fact that we found EC to be insensitive to context manipulations is in line with the findings of Baeyens and colleagues (Baeyens et al., 1996, 1998), who also failed to obtain occasion setting in EC. Our results contrast, however, with those of Hardwick and Lipp (2000) and Rydell and Gawronski (2009). Conflicting results are widespread in the EC literature (for reviews, see De Houwer, Thomas, & Baeyens, 2001; De Houwer et al., 2005; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, in press). Just as there is controversy about the context-sensitivity of EC, there are intense debates about whether or not EC is resistant to extinction (e.g., Lipp, Oughton, & Lelievre, 2003; Vansteenwegen et al., 2006), depends on contingency awareness (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Walther & Nagengast, 2006), is sensitive to cue competition (e.g., Beckers, de Vicq, & Baeyens, 2009; Lipp, Neumann, & Mason, 2001), etc. Hitherto, there is no clear explanation for the large body of conflicting findings in the EC literature. On the one hand, methodological and procedural factors might contribute to discrepant findings. The difference in findings between our study and the study of Hardwick and Lipp (2000) might, for instance, result from the different measures that were used to assess participants' evaluations. As already mentioned, there is evidence that the startle response does not provide a good index of EC because it can also be affected by factors other than stimulus valence (e.g., Cuthbert, Bradley, & Lang, 1996). Similarly, a close comparison with the results of Rydell and Gawronski (2009) is difficult as these authors only measured participants' evaluations indirectly. Our study also included an indirect measure, but, as explained above, it is not clear whether the data it provided are valid. Perhaps our rating results were biased by demand effects (but see earlier for arguments against this idea) and therefore differed from the observations of Rydell and Gawronski (2009) and/or Hardwick and Lipp (2000).

Some authors, on the other hand, have argued that the opposing results in the EC literature might reflect the operation of different processes. De Houwer et al. (2005; De Houwer, 2007), for instance, suggest that in some studies EC effects might depend on the automatic formation of associations while in other studies the effects might be based on the formation of conscious propositional knowledge about the CS-US relation(s) (or even other processes). An explanation for the difference between our findings and those of Rydell and Gawronski (2009) and Hardwick and Lipp (2000) might then, for instance, be that modulation occurs only for EC effects that rely on the formation of conscious propositional knowledge about the conditional CS-US relations. In our study, EC effects might have been driven by a more simple learning process (e.g., the automatic formation of associations governed by a rudimentary Hebbian learning rule, see also Baeyens et al., 1998) that is not sensitive for context information. Note, however, that this explanation is post-hoc and that there are equally good reasons to assume that propositional processes might have played a role in the current study. The post-conditioning questionnaire data, for instance, revealed that the vast majority of the participants were aware of the context manipulation and the conditional contingencies. Participants' evaluations also changed very quickly throughout the experiment (already after 3 trials per CS: see Figure 7). The fact that there was room for propositional reasoning in the present experiment does not imply, however, that participants' preferences were rule-based. Also in this context, a post-experimental questionnaire on participants' thoughts and reasons behind their ratings might have been enlightening. It is clear, however, that determining what processes underlie particular EC effects is not an easy task (for a more elaborate discussion on this topic see De Houwer, 2007, in press). In sum, the aforementioned considerations indicate that research on EC is actually still in its early stages and that more experiments are needed before the pieces of the EC puzzle can even start to fall in place.

A remarkable finding in the contingency recall data was that participants of the AAA but not ABA groups remembered fewer contingencies of the acquisition phase than of the CC phase. This finding is most likely an artifact resulting from the way in which contingency recall was assessed. In the contingency recall questionnaire, we asked participants to indicate for each CS with which US it had been paired during the '*first*' and the '*second part*' of the conditioning phase. For the participants of the ABA groups, the background screens made clear which parts of the conditioning procedure we were referring to (i.e., part 1 = context A, part 2 = context B). For the participants of the AAA groups, on the other hand, it might not have been clear what we meant with these terms as for them there was no context switch in between acquisition and CC. These phases were moreover not equal in length: the CC phase was twice as long as the acquisition phase. Therefore, participants in the AAA groups might have been confused about what constituted the first part of the conditioning procedure, which could explain the fewer correct answers for the acquisition phase in these groups.

Finally, besides theoretical importance, the present findings potentially have important implications for preference change in applied settings. The rating data corroborate the findings of Experiment 3 in suggesting that CC forms a promising technique for altering conditioned preferences. Our data further suggest that these altered preferences are not (necessarily) context-bound. For clinical practice, this finding is reassuring as it suggests that new evaluations that patients acquire through CC can outlast the therapeutic context. More research is needed, however, to confirm this conclusion.

References

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation, 27*(2), 200-241.

Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither extended sequential nor simultaneous feature positive training result in modulation of evaluative flavor-flavor conditioning in humans. *Appetite, 31*(2), 185-204.

Beckers, T., de Vicq, P., & Baeyens, F. (2009). Evaluative conditioning is insensitive to blocking. *Psychologica Belgica, 49*, 41-57.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*(10), 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*, 485-494.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33*, 103-111.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *The Spanish Journal of Psychology, 10*, 230-241.

De Houwer, J. (in press). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds.), *Applications of learning and conditioning*. Oxford, UK: Oxford University Press.

De Houwer, J., Baeyens, F., & Field, A.P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19,* 161-174.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*, 853-869.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*(2), 245-268.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11,* 549-554.

Fonteyne, R., Vervliet, B., Hermans, D., Baeyens, F., & Vansteenwegen, D. (2009). Reducing chronic anxiety by making the threatening event predictable: An experimental approach. *Behaviour Research and Therapy, 47*(10), 830-839.

Gawronski, B., Walther, E., & Blank, H. (2005). Cognitive consistency and the formation of interpersonal attitudes: Cognitive balance affects the encoding of social information. *Journal of Experimental Social Psychology, 41*(6), 618-626.

Hardwick, S. A., & Lipp, O. V. (2000). Modulation of affective learning: An occasion for evaluative conditioning? *Learning and Motivation, 31*(3), 251-271.

Hermans, D., & De Houwer, J. (1994). Affective and subjective familiarity ratings of 740 Dutch words. *Psychologica Belgica, 34*(2-3), 115-139.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (in press). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*.

Huijding, J., & de Jong, P. J. (2005). *Changes of automatic and self-reported spider phobia-related affective associations following treatment and the prediction of symptom return*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen.

Holland, P. C. (1992). Occasion setting in pavlovian conditioning. *Psychology of Learning and Motivation - Advances in Research and Theory, 28*, 69-125.

Lipp, O. V., Neumann, D. L., Mason, V. (2001). Stimulus competition in affective and relational learning. *Learning and Motivation, 32*, 306-331.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative learning in human pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology - Learning Memory and Cognition, 33*(1), 130-144.

Rosas, J. M., Vila, N. J., Lugo, M., & Lopez, L. (2001). Combined effect of context change and retention interval on interference in causality judgments. *Journal of Experimental Psychology - Animal Behavior Processes, 27*(2), 153-164.

Rydell, R. J., & Gawronski, B. (2009). I like you, I like you not: Understanding the formation of context-dependent automatic attitudes. *Cognition & Emotion, 23*(6), 1118-1152.

Swartzentruber, D. (1995). Modulatory mechanisms in Pavlovian conditioning. *Animal Learning and Behavior, 23*, 123-143.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(1), 71-79.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquisition context. *Behaviour Research and Therapy, 43*(3), 323-336.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 454-459.

experiment 5

The role of negative affective valence in return of fear: A study in mice

Return of fear after successful exposure is a common finding and constitutes a challenge for clinical practice and fear research. According to contemporary learning theories, fear can be acquired through an associative learning process. In a fear conditioning procedure a neutral stimulus (the Conditioned Stimulus or CS) comes to evoke fear after it was repeatedly paired with an aversive stimulus (the Unconditioned Stimulus or US). From this perspective extinction - a decrease in conditioned responding to the CS due to repeated presentations of the CS alone - can be viewed as an experimental model for exposure therapy. Return of conditioned responses after extinction then can be seen as a model for relapse after treatment. Experimental research suggests that during a fear acquisition procedure the reinforced CS (CS+) does not only become a valid predictor for the US but also acquires a negative connotation that is resistant to extinction. This is in line with the clinical observation that even after avoidance behavior has been drastically reduced through exposure, spider fearful individuals continue to consider spiders as negative animals (Marks, 1987). This remaining negative valence after extinction/exposure might function as an affective-motivational source for return of fear. Procedures that alter the valence of the fear eliciting stimulus after extinction/exposure (e.g., counterconditioning procedures) might then be effective in diminishing return of fear. In this chapter, we will give an overview of the research literature on the role of affective valence in (return of) fear. In addition, we discuss results of an animal (mice) study that was designed to investigate the impact of a counterconditioning procedure after extinction on return of conditioned responding.

Published as: Kerkhof, I., Vansteenwegen, D., Beckers, T., Dirikx, T., Baeyens, F., D'Hooge, R., & Hermans, D. (2009). The role of negative affective valence in return of fear. In H. D. Friedman & P. K. Revera (Eds.), *Abnormal psychology: New research* (pp. 121-138). New York: Nova Science Publishers, Inc.

Introduction

The possible role of negative affective stimulus valence in return of fear

Epidemiological research shows that about 30 percent of the general population will develop an anxiety disorder at some point in life (Kessler, Koretz, Merikangas, & Wang, 2004). Not surprisingly, a lot of research is invested in unveiling the processes that are responsible for the origin and maintenance of fear, as well as in the principles that constitute the core of successful treatment. In this context, there have been a lot of developments during the last four decades. The general efficacy of exposure-based treatments for anxiety disorders is now irrefutable. For some anxiety disorders like simple phobia, treatment success is so high that it is almost unequalled in our health services in general (e.g., Öst, 1989). Also for disorders like panic disorder, agoraphobia, post-traumatic stress disorder and social phobia, treatment outcome of exposure-based treatments is generally good (e.g., van Balkom et al., 1997). These positive developments have led to a shift in research goals. Because of the successes at the curative level, more attention has been devoted to primary and secondary prevention. Concerning the latter (secondary or relapse prevention) it is important to acknowledge that a number of patients - in spite of apparently successful treatment - experience a return of symptoms of fear and anxiety. For some this is a source for complete relapse. This return of symptoms of fear and anxiety after successful treatment is known as 'return of fear' (ROF, Rachman, 1989) and constitutes a challenge for clinical practice and fear research.

Contemporary models of human classical conditioning and more specifically of fear conditioning, provide a rich conceptual framework for the understanding of the etiology, maintenance and treatment of human fears and phobias (Craske, Hermans, & Vansteenwegen, 2006). The essence of these models is that they view classical conditioning as the acquisition of associations between memory representations. For instance, being repeatedly confronted with a panic attack (referred to as the Unconditioned Stimulus or US) in the context of a supermarket (referred to as the Conditioned Stimulus or CS), can lead to the formation of an association between the memory representations of both stimuli/events. Later confrontation with the CS will activate the representation of the US, as well as the fear that is associated with it. From this perspective, exposure therapy can be viewed as the clinical analogue of extinction (i.e., unreinforced presentations of the CS after acquisition that result in a decrease in conditioned responding). ROF after treatment then can be seen as the clinical analogue of return of conditioned responses after extinction.

In the context of the current chapter, two recent insights from fear research are important. First, during the last decennia it became clear that extinction does not reflect an ‘unlearning’ of the underlying CS-US association (Bouton, 2004). Instead, extinction reflects new (inhibitory) learning, leaving the initial CS-US association intact. This original association can be activated by several ‘post-extinction events’, resulting in a return of the extinguished responses. Examples of such post-extinction events are the mere passage of time (‘spontaneous recovery’), a change of context (‘renewal’) or the presentation of US-only trials after extinction (‘reinstatement’).

A second relevant insight is that during a fear acquisition procedure the meaning of the CS alters in two important ways. The CS not only becomes a valid predictor for the US (which results in its presentation being accompanied by increasing levels of fear as the CS activates the expectation of US occurrence), but also acquires a negative connotation through evaluative conditioning (De Houwer, Thomas, & Baeyens, 2001). Hence, two types of learning co-occur in a fear conditioning procedure: expectancy or predictive learning (i.e., participants learn that the CS predicts US occurrence, they expect the US to follow when they see the CS) and evaluative learning (Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2002). Importantly, several studies found the acquired negative connotation of the CS - in contrast to its acquired signal value - to be (rather) unaffected by an extinction procedure (Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Hermans et al., 2002; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006). Significant levels of negative valence can still be observed for the CS, even if extinction reduces the US-expectancy to pre-acquisition level (Hermans et al., 2002; Vansteenwegen et al., 2006). If we translate this to a clinical situation, these findings suggest that to the extent that the acquisition of clinical anxiety (e.g., agoraphobia) is based on the contingent presentation of an originally neutral stimulus (e.g., a supermarket) and an aversive event (e.g., a panic attack), this might not only result in the supermarket becoming a valid predictor for a panic attack, but also in an affective shift for supermarkets. Moreover, the findings concerning the relative resistance to extinction of this acquired evaluative meaning suggest that exposure treatment for clinical fear might successfully reduce the expectancy component in clinical fear (and therefore might lead to diminished fear reactions), but might leave the acquired affective meaning relatively unaltered. In our example, this would mean that the patient no longer expects a panic attack when entering a large supermarket, and after successful exposure treatment again frequents these superstores, but still somehow dislikes them. Clinical experience indicates that this differential outcome can indeed be observed (Marks, 1987).

In addition to the fact that the CS-US association can survive extinction and hence form the basis for ROF, we consider the evaluatively conditioned, but extinction-resistant negative valence of the CS as a second possible source for the re-emergence of the original phobic fear. There are several ways in which residual negative associations could increase the risk of relapse. In general, negative stimuli are more easily associated with aversive outcomes than are neutral or positive stimuli (e.g., Hamm, Vaitl, & Lang, 1989). Moreover, negative valence is associated with action tendencies of escape and avoidance (e.g., Chen & Bargh, 1999). Encounters with the previously phobic object are therefore more likely to reinstate phobic fear if the object elicits negative affect then when it would elicit neutral or positive affect. An alternative route may follow from the emotion theory proposed by Lang (1995). According to Lang, all emotions can be situated in a two-dimensional space, with affective valence (positive/negative) and arousal (high/low) as crucial dimensions. These dimensions interact to constitute different emotions. Against this background, fear can be considered as an emotion characterized by a combination of negative valence and high levels of arousal. Extinction/exposure will lead to a significant decrease in arousal, but leaves the negative stimulus valence intact (cf. Hermans et al., 2002). Following this, fear for the CS may re-emerge relatively easily if this still negatively valenced CS is encountered in an arousing context and negative valence and a higher level of arousal are thus recombined.

A reinstatement procedure, for example, that comprises presenting US-only trials after extinction, could be considered as a procedure that essentially endows the context with a higher level of arousal. Importantly, re-emergence of fear might even occur if the experienced arousal is in fact unrelated to the US. In our example this would imply that if after his/her treatment the patient enters a supermarket that is still endowed with negative valence, and he/she is aroused for reasons that are unrelated to this situation (e.g., bad news, too much coffee), this combination might lead to enhanced fear or even a new panic attack.

The findings concerning return of conditioned fear responses after extinction and extinction-resistant stimulus valence are both theoretically and clinically relevant. They suggest that simple extinction/exposure might not be sufficient and raise the possibility that procedures that alter the valence of the fear-eliciting stimulus could function as an adjunct to mere CS-only presentations and might be effective in diminishing ROF. Up till now, however, research on the role of negative affective stimulus valence in ROF is rather sparse, both in humans and animals. In line with our hypothesis two laboratory studies on reinstatement in humans found the remaining negative valence of the reinforced CS (CS+) after extinction to be predictive for the return of fear for this CS+ after reinstatement (Dirikx et al., 2004; Hermans et al., 2005). The more negative the CS+, the more reinstatement was observed. Further evidence comes from a study in spider phobics by Huijding and de Jong (2005), who found residual negative associations after exposure treatment to be predictive for symptom return (i.e., return of overt avoidance behavior) at two-month follow-up. The findings of these studies are, however, post-hoc and correlational. In the present chapter a study in laboratory rodents will be presented in which the role of negative affective valence in ROF was investigated *experimentally* by manipulating the valence of the CS+ after extinction through a counterconditioning (CC) procedure. Before we go into this study, we will first take a look at the literature on CC. In a typical CC procedure a CS is paired with a US of one affective sign (positive/negative) in a first phase and with a US of the opposite sign (negative/positive) in a subsequent phase (generally referred to as the CC phase). As explained above, two outcomes can co-occur when a CS is paired with a valenced US: expectancy learning (the CS acquires signal value, the US is expected to follow after the CS) and evaluative learning (the valence of the CS shifts in the direction of the valence of the US). Because in a CC phase a CS is paired with a new US, CC might affect both types of learning. In the next section we will look at studies that investigated the impact of CC on indices of evaluative learning. In human studies typical indices of evaluative learning are valence ratings and indirect reaction time tasks (e.g., the Affective Priming Paradigm of Fazio, Sanbonmatsu, Powell, & Kardes, 1986 or Greenwald, McGhee, & Schwartz's (1998) Implicit Association Test). In animal studies, evaluative learning is generally inferred from preference choice tests (e.g., two-bottle preference tests, Scalera, 2000). Subsequently, we will focus on CC in relation to expectancy learning. Two indices commonly used to assess expectancy learning in humans are electrodermal activity and US-expectancy ratings. In animals, expectancy learning is often assessed with a suppression index (see further) or by scoring US-specific anticipatory behavior (e.g., defensive or appetitive responses).

The influence of counterconditioning on indices of evaluative learning

As mentioned in the introduction, several studies provide evidence for evaluative learning to be resistant to extinction (see De Houwer, Baeyens, & Field, 2005 for an overview). It appears to be the case that once a stimulus has acquired a valence as the result of being paired with a liked or disliked stimulus, this acquired valence cannot readily be changed by repeatedly presenting this stimulus on its own. In a CC procedure a CS is not presented without US, but is paired with a new oppositely valenced US. Research on evaluative conditioning suggests that the new CS-US pairings in the second phase of a CC experiment have an impact on the evaluative value the CS acquired in the first phase. We will now give an overview of this research.

Several studies in humans looked at the effect of a CC procedure on evaluative learning. Baeyens, Eelen, Van den Berg, and Crombez (1989) were the first to demonstrate that the conditioned valence of a CS can be eliminated and even changed into an opposite evaluation by a CC procedure. Baeyens et al. (1989) conducted a picture-picture evaluative conditioning study that started with a baseline measurement of valence followed by an acquisition phase in which neutral faces (CSs) were paired with liked, disliked or other neutral faces (USs). Following this acquisition phase, some of the CSs were involved in a CC treatment, which entailed that those CSs were paired with a new US of a valence opposite to the valence of the US with which they were previously paired. Other CSs were presented on their own during the second phase (extinction treatment), whereas still other CSs were not presented during the second phase (control treatment). After the second phase, participants were required to rerate the valence of all CSs (test). Results showed that the liking of extinction CSs and control CSs increased from baseline to test if they were paired with a positive US during acquisition, but decreased if they were paired with a negative US. Hence, the extinction procedure did not abolish the influence of the acquisition contingencies. In contrast, the liking of CSs that were submitted to a CC treatment was the same on baseline and test, with a trend towards a shift in valence in the direction of the US with which these CSs were paired in the second phase. The influence of the acquisition contingencies was thus eliminated by the CC treatment. Since this first study, successful CC effects were obtained in several other human studies using different evaluative conditioning paradigms (e.g., flavor-taste conditioning, Baeyens, Crombez, De Houwer, & Eelen, 1996; picture-picture conditioning, Lipp & Purkis, 2006), and different direct and indirect measures (e.g., ratings, Baeyens et al., 1996; the Implicit Association Test, Rydell & McConnell, 2006; Rydell, McConnell, Strain, Claypool, & Hugenberg, 2007; the Initial-Preference Task, Dijksterhuis 2004). A number of studies even succeeded in changing pre-existing (rather than recently acquired) attitudes through CC (e.g., the negative attitude towards spiders of spider phobics, Eifert, Craill, Carey, & O'Conner, 1988; implicit self-esteem, Dijksterhuis, 2004; racial prejudice, Olson & Fazio, 2006).

We are unaware of any animal study that looked at the impact of CC on evaluative learning (i.e., on the valence of the CS). Most likely this is related to the fact that adequate valence measures for animals are difficult to develop.

The influence of counterconditioning on indices of expectancy learning

In this section we will look at studies that investigated the impact of CC on indices of expectancy learning. We are unaware of any experimental studies in humans that looked at the influence on expectancy learning of pairing a CS with differently valenced USs in successive phases. Several studies in rats, however, have addressed this issue and looked at the impact of presenting CS-food pairings after CS-shock pairings or the reverse (e.g., Bouton & Peck, 1992; Brooks, Hale, Nelson, & Bouton, 1995; Delprato & Jackson, 1973; Peck & Bouton, 1990). In all these studies, the CS was found to evoke responding appropriate to the second US by the end of the second phase. Bouton and Peck (1992), for example, observed strong freezing responses (which are natural defensive responses of rats indicative of fear) towards a tone that was initially paired with shock. These freezing responses were, however, gradually replaced by head-jerk responses (which are appetitive responses a rat makes to auditory stimuli that have been associated with food) in a second phase in which the tone was paired with food. Hence, CC resulted in a shift in expectancy learning: performance corresponding to the second association replaced performance to the first. An important note to make here is that even though CC was effective in establishing a new conditioned response towards the CS, this does not imply that the initially learned CS-shock association was destroyed. Similar to what was found for extinction, several studies found CC to be incapable of destroying the previously acquired association. In several studies by Bouton and colleagues (Bouton & Peck, 1992; Brooks et al., 1995; Peck & Bouton, 1990) rats received CS-shock pairings followed by CS-food pairings. Even though appetitive responses were observed at the end of the second phase, a return of the original fear response towards the CS was observed after the mere passage of time (i.e., spontaneous recovery, Bouton & Peck, 1992), after a change of context (i.e., renewal, Peck & Bouton, 1990) and after the presentation of a number of unsignaled shocks (i.e., reinstatement, Brooks et al., 1995). According to Bouton (2002, 2004), both extinction and CC reflect new context-dependent learning rather than unlearning. Hence, ROF can still occur after CC.

Counterconditioning as a technique to reduce return of fear?

Based on the aforementioned research, we can conclude that whereas extinction only seems to have an influence on expectancy learning, CC seems to be capable of influencing both expectancy learning and evaluative learning. Previous studies further suggest that ROF can still be observed after a CC procedure. These studies, however, do not exclude the possibility that ROF can be *reduced* by CC as in those studies the amount of ROF after CC was never compared to a control group that received no CC trials. Therefore, in the present mice study, we wanted to examine if CC can have a beneficial effect on ROF when a group that undergoes extinction and CC is compared to a group that only receives extinction trials. We expected to find less ROF after CC, based on the idea that the remaining negative valence of the CS+ after extinction can form an affective-motivational source for ROF and that this negative connotation can be changed through CC (and not through extinction).

Despite the fact that evaluative CC techniques are sometimes applied in clinical practice (e.g., Korrelboom, van der Gaag, Hendriks, Huijbrechts, & Berretty, 2008), experimental research on the effectiveness of applying such techniques to reduce (return of) fear is scarce. There is only one experimental study we are aware of that has looked at the impact of changing valence through CC on fear. De Jong, Vorage and Van den Hout (2000) compared a group of spider phobics that received a standard one-session exposure treatment to a group that received an equally long treatment that consisted of exposure and CC. In the CC group, tasty food-items were used during the standard exposure exercises and the participants' favorite music was played. In contrast to what was expected, no difference in treatment efficacy was observed between both groups when tested immediately after treatment and at 1 year follow-up. Moreover, CC was not found to be more effective in altering the affective valence of spiders than regular exposure treatment; both procedures resulted in a significant reduction in the negative evaluation of spiders. There are, however, some important limitations to the De Jong and al. (2000) study. A first limitation is that the standard exposure sessions probably included several ingredients that may have helped to undermine the negative affective valence of spiders and can be conceptualized as forms of CC (e.g., accepting expressions by the therapist towards spiders, information about spiders that portrays them as tender, fragile and timid animals). A second remark is that both procedures altered the affective evaluations of spiders in a positive direction, but did not lead to the abolishment of spiders' negative valence; an important amount of negative valence remained after both treatments. Perhaps the CC procedure was too brief in duration or not of sufficient strength to neutralize the spiders' negative valence.

In the present study we wanted to further investigate the possible impact on ROF of a CC procedure aimed at changing the negative valence of the CS+. Therefore, a reinstatement study was set up in laboratory rodents (mice). In previous research in our lab (Dirikx et al., 2007), ROF after reinstatement was assessed in mice using a differential conditioned suppression paradigm. In a first phase, one CS was consistently paired with a US (foot shock) while another CS was not, resulting in selective suppression of previously trained instrumental behavior during the CS+. After the extinction phase, half of the animals (reinstatement group) received unsignaled USs while the other half (control group) did not. A differential return of conditioned responding was observed in the reinstatement group only. In the present study the same differential conditioned suppression paradigm was used, but with an extra phase added in between the extinction and reinstatement phase. During this phase, half of the mice (counterconditioning group; CC group) received CC trials during which the CS+ was paired with food pellets, while the other half (further extinction group; FE group) received additional extinction trials. We expected to find less reinstatement (i.e., ROF) for the CS+ in the CC group as compared to the FE group.

Method

Subjects

The subjects were 49 experimentally naive female C57BL mice from Elevage Janvier, Le Genest St. Isle, France. All animals were kept under standard laboratory conditions on a 12h/12h dark-light schedule, with constant room temperature and humidity, standard lab chow and water ad libitum. The animals were between 91 and 105 days old at the start of shaping. The animals were food-deprived prior to the start of the study and maintained on 80 to 90% of their initial body weight.

Apparatus and stimuli

Eight operant soundproof chambers (Coulbourn Instruments, Allentown, PA) were used for shaping and classical conditioning. The conditioned stimuli were 30 s tones (4000 Hz or 1000 Hz). Due to a procedural error the assignment of tones to CS-type was not fully counterbalanced across subjects; for 31 subjects the 1000 Hz tone was used as the CS+ and the 4000 Hz tone as the CS-, while for the remaining 18 subjects the reverse was true. A 200 ms foot shock (0.2 mA) was used as the US. Sucrose pellets (Noyes Precision Pellets, Formula F from Research Diets Inc., New Brunswick, USA) were used for shaping and classical conditioning. During the CC phase grain pellets (Noyes Precision Pellets, Formula A/I) were used because preference tests administered before the start of the experiment had indicated that the mice preferred these pellets over sucrose pellets.[33] All stimulus presentations and response registrations were programmed with Graphic State 3.0 software (Coulbourn Instruments, Allentown, PA).

Procedure

In the present study fear conditioning was assessed using a conditioned suppression paradigm. In a typical conditioned suppression paradigm animals are first shaped to emit an operant response (e.g., lever pressing, nose poking) to obtain a positive reinforcer (e.g., food pellets) until a stable response rate is reached. Following this pre-training a Pavlovian conditioning procedure (e.g., a tone followed by a shock) is superimposed on this ongoing operant responding. The conditioned effect is a disruption of the operant responding. If the animal is scared of the CS, it will probably freeze and emit fewer operant responses during the CS than in the immediately preceding pre-CS interval. Hence, learning of the Pavlovian associations can be derived from the amount of suppression of the operant response. This amount of suppression is measured

[33] During the preference tests, mice had access to two different kinds of pellets (20 g of each kind) in their home cage for 14 hours. After this period, the intake weight for each type of pellet was calculated by subtracting the weight of the remaining pellets from the original weight. Three types of pellets were compared: standard purified pellets (Noyes Precision Pellets, Formula P), sucrose pellets (Formula F) and grain pellets (Formula A/I). A significant difference in intake weight was only observed between sucrose and grain pellets with mice eating more grain pellets (M = 14 g) than sucrose pellets (M = 10 g), $t(7) = -10.40$, $p < .0001$.

by calculating a suppression ratio (A/A+B) with A and B representing response rates during the CS and during equal periods of time immediately before CS onset, respectively. A suppression ratio of .50 indicates an equal response rate in the presence of the CS versus in the absence of the CS, thus complete lack of suppression (i.e., no fear). Complete suppression of responding during the CS (i.e., strong fear) is represented by a ratio of 0. In the current study suppression of previously trained nose poking behavior was used as an index of fear conditioning.

The experiment consisted of six phases: shaping of instrumental responding, acquisition, extinction, CC or further extinction, reinstatement and test of conditioned emotional responding. The first three phases and the last phase were identical for all animals. The experiment took 116 days in total. One training or test session was scheduled per day and every session took approximately 30 minutes. The 49 mice were housed in seven groups of six animals and one group of seven animals that were initially tested together. On day 18 of the extinction phase, the mice were assigned to one of the four experimental groups. Thirteen mice (8 with the 1000 Hz tone and 5 with the 4000 Hz tone as CS+) were assigned to the CC-reinstatement group, 12 (1000 Hz: 8, 4000 Hz: 4) were assigned to the CC-control group, 12 (1000 Hz: 7, 4000 Hz: 5) to the FE-reinstatement group and 12 (1000 Hz: 8, 4000 Hz: 4) to the FE-control group. The mice were also rehoused so that reinstatement and control animals were in separate home cages. We did this to avoid that control mice would notice fear reactions in the mice of the reinstatement groups during and after the reinstatement phase.

Shaping. In an operant shaping procedure, all mice were gradually trained to use a nose poke device to obtain sucrose pellets. During the first six sessions, a CRF schedule (Continuous Reinforcement) was applied, implying that every nose poke was reinforced by delivery of a food pellet. In the first three CRF sessions additional food pellets were given every two minutes regardless of nose poking behavior to alert the animals to the very possibility of food pellet delivery, and to demonstrate the contingency between the mechanical sound of the food pellet dispenser and food pellet delivery. Subsequently, two FR5 (Fixed Ratio, 5 nose pokes) and two FR10 sessions were programmed, in which every fifth or tenth nose poke was reinforced, respectively. The FR sessions were followed by nine VR10 (Variable Ratio, 10 nose pokes) sessions, in which reinforcement occurred after a mean of 10 nose pokes (range 4-16). Subsequently, the mice received five sessions of VI 30 s (Variable Interval with a mean of 30 s) training during which a nose poke was reinforced after mean intervals of 30 s (range 15-45). Additional shaping sessions followed after these VI 30 s sessions because shaping was interrupted for a week and because we wanted to achieve high and stable response rates. These additional sessions consisted of two FR10 sessions, 10 VR10 and 11 VI 30 s sessions.

Habituation. Before the start of acquisition, one habituation session was scheduled. During this session both CSs were presented six times without US, superimposed on the VI 30 s schedule (i.e., superimposed on the ongoing operant nose poke response). Suppression of this nose poking response constituted the index of fear for the CSs in this and the following phases.

Acquisition. During acquisition a Pavlovian fear conditioning procedure was superimposed on the VI 30 s reinforcement schedule. One 30 s tone (CS+) co-terminated with a 200 ms foot shock. The second CS (CS-) never co-terminated with foot shock. Each of the 17 acquisition sessions contained 6 presentations of every CS with a mean intertrial interval (ITI) of approximately 80 s.

Extinction. None of the tone CSs was followed by the US during the extinction phase. Otherwise presentation parameters were the same as during acquisition. There were 20 extinction sessions in total.

Counterconditioning or further extinction. After the extinction phase, half of the mice (CC group) received 17 CC sessions during which the CS+ was followed by delivery of a grain pellet, while the CS- was not. The other half of the mice (FE group) received further extinction sessions. For both groups the nose poke devices were removed from the chambers. We did this because pellets could only be delivered through one feeder hole in the chamber and we wanted to make sure that the mice of the CC group would notice the new contingencies (i.e., pellet deliverance is related to the presentation of the CS+ and not merely to nose poking like in the previous phases). For a similar reason (i.e., because the mice would learn that pellet deliverance is only related to CS+ and not to CS- presentations), the CS+ and CS- were presented blocked in this phase of the experiment. Each 30 minutes session consisted of 15 minutes during which the CS+ was presented six times with an ITI of approximately 100 s and 15 minutes during which the CS- was presented six times with an ITI of approximately 100 s.

Nose poke retraining. After the CC / FE phase, the nose poke devices were placed back in the chambers and all mice were retrained to nose poke at a stable rate. The mice received one FR5 session, five FR10 sessions and six VI 30 s sessions.

Reinstatement. During the reinstatement session, half of the mice of the CC group (CC-reinstatement group) and half of the mice of the FE group (FE-reinstatement group) received four unsignaled USs. The USs were administered at approximately 6 min, 10 min, 18 min and 23 min after the onset of the session. In the control groups (CC-control group and FE-control group) no USs were presented during the reinstatement session.

Test. Twenty-four hours after the reinstatement treatment a test session took place. During this test session both CSs were tested six times, without reinforcement.

Results

Figure 9 provides an overview of the mean suppression ratios for the CS+ and CS- during the habituation, acquisition and extinction phase.

Scoring and statistics

The degree of fear conditioning was assessed by determining if the normally ongoing nose poking behavior was disrupted by the CS presentations. To measure suppression to the tones, for each session a suppression ratio (A/A+B) was calculated with A and B representing response rates during the CSs and during equal periods of time immediately before CS onset, respectively. As explained earlier, a suppression ratio of .50 indicates complete lack of suppression and thus no fear, while complete suppression of responding (i.e., strong fear) during the CS is represented by a ratio of 0. An alpha level of .05 was used for all statistical tests.

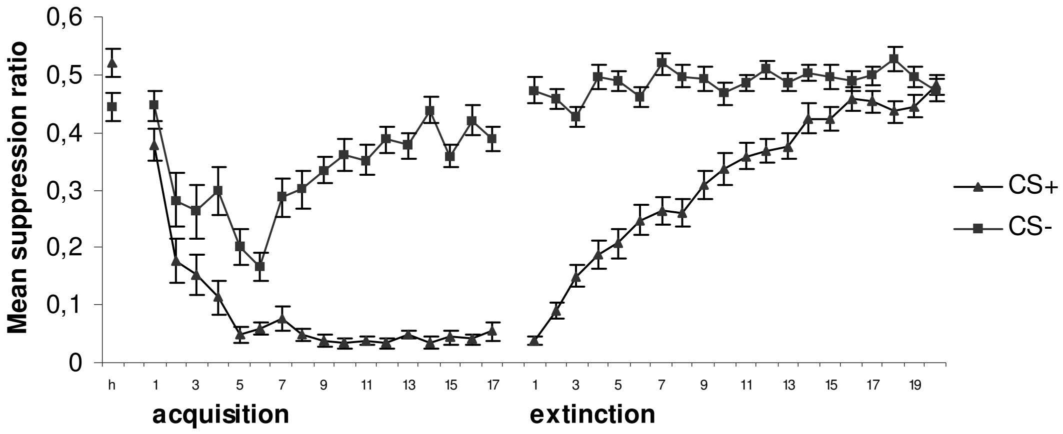


Figure 9. Mean suppression ratios for the total group during the habituation (h), acquisition (1-17) and extinction (1-20) sessions. Lower values represent stronger conditioned responding. Error bars represent standard errors of means.

Habituation and acquisition

We expected to observe a similar low or absent amount of fear for both CSs during the habituation phase and at the very beginning of acquisition as both tones are initially new and neutral for the mice. An increase in fear throughout the acquisition phase was expected for the reinforced CS, while little or no change in fear was expected for the CS-. In terms of suppression ratios this implies that we expected high suppression ratios during habituation and at the beginning of acquisition. Throughout acquisition, we expected to observe a decrease in the suppression ratio for the CS+ and little or no change in the suppression ratio for the CS-. In addition, we expected to find no differences between the four experimental groups as all groups received the same stimulus presentations and contingencies in the habituation and acquisition phase.

As shown in Figure 9, the CS+ and the CS- elicited similar low amounts of suppression (i.e., low amounts of conditioned fear) during the habituation phase. The 2 (Condition: reinstatement vs. control) x 2 (Treatment: CC vs. FE) x 2 (CS-type: CS+ vs. CS-) analysis of variance (ANOVA) showed neither a main effect of CS-type, $F(1, 45) = 3.28$, $MSE = .04$, $p = .08$, nor a significant interaction with condition or treatment, both with $F < 1$. The mean suppression ratios for the CS+ and CS- were .52 and .44, respectively. The absence of a main effect of CS-type indicates that, in line with what we expected, the CS+ and CS- did not differ in the extent to which they elicited suppression (i.e., fear) before the start of the acquisition phase. The absence of an interaction with condition and treatment signifies that this effect is the same in the different experimental groups.

In comparison with the first acquisition session, suppression to the CS+ was much stronger by the end of acquisition, while for the CS- little change in suppression was observed. The mean suppression ratios for all 17 acquisition sessions are depicted in Figure 9. A 2 x 2 x 2 x 2 ANOVA with condition and treatment as between-subjects variables and CS-type and moment (first acquisition session vs. last acquisition session) as within-subjects variables confirmed the differential acquisition of conditioned suppression. The moment by CS-type interaction proved to be highly significant, $F(1, 44) = 37.50$, $MSE = .02$, $p < .001$, clearly showing differential acquisition of fear. There was a significant increase in suppression and thus in conditioned fear for the CS+, $F(1, 44) = 120.73$, $MSE = .02$, $p < .001$, with mean suppression ratios of .38 and .05 during the first and last acquisition session, respectively. For the CS-, on the other hand, no significant change in suppression (i.e., in conditioned fear) was observed from the first to the last acquisition session, $F(1, 44) = 3.32$, $MSE = .02$, $p = .08$, with mean suppression ratios of .45 and .39, respectively. None of these effects interacted with the condition or treatment variables. A similar pattern of data was thus observed in the different experimental groups.

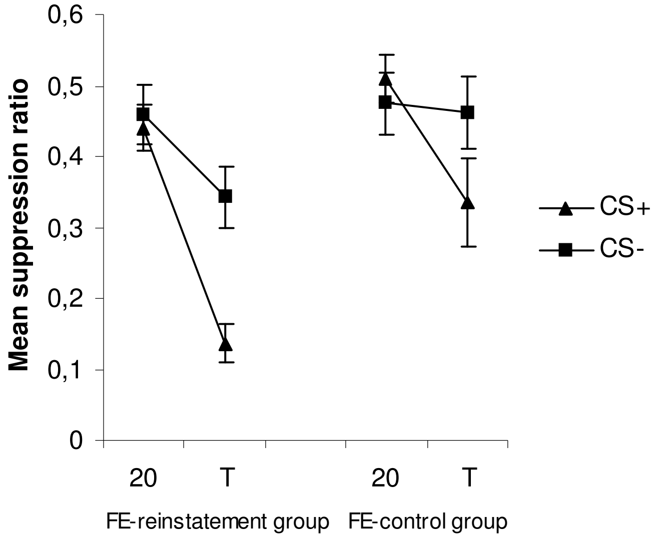
In conclusion we can say that the results of the habituation and acquisition phase were fully in line with expectations. Similar low amounts of conditioned fear were observed for the CS+ and CS- during habituation and at the beginning of acquisition. Throughout the acquisition phase the mice gradually acquired the differential contingencies between the CSs and the US and showed an increase in conditioned fear for the CS+ but not for the CS-.

Extinction

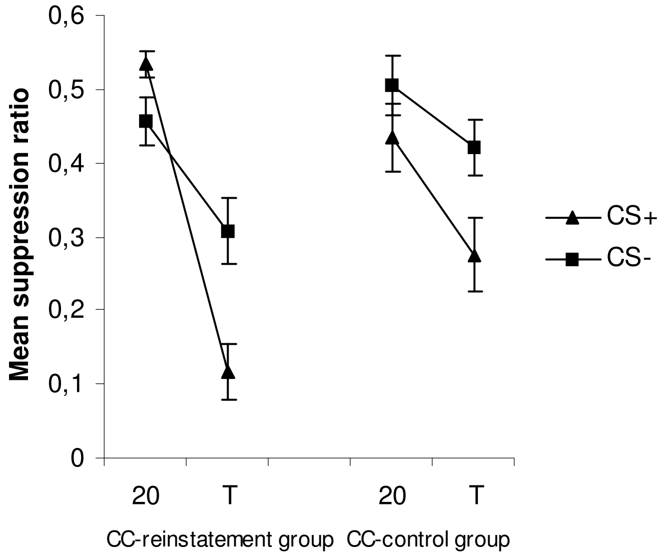
During the extinction phase, we expected to observe a gradual diminishment of the conditioned fear for the CS+. By the end of extinction we expected the CS+ to elicit similar low amounts of fear as the CS-. In terms of suppression ratios we thus expected to see an increase in the suppression ratio for the CS+ during extinction and high suppression ratios for both the CS+ and CS- at the end of extinction. Also for the extinction phase, no differences were expected between the four experimental groups as the extinction procedure was the same in all groups.

As shown in Figure 9, as predicted, the mean suppression ratio for the CS+ increased substantially from the end of acquisition to the end of the extinction phase. The difference in suppression elicited by the CS+ and CS- by the end of extinction was much smaller than the difference observed at the end of the acquisition phase. To test for extinction, the mean suppression ratios of the last acquisition session and the last extinction session were compared in a 2 (Condition) x 2 (Treatment) x 2 (CS-type) x 2 (Moment) ANOVA. As expected, the moment by CS-type interaction was highly significant, $F(1, 45) = 89.75$, $MSE = .02$, $p < .001$. A significant decrease in suppression was observed both for the CS+ and CS-, with respectively $F(1, 45) = 299$, $MSE = .01$, $p < .001$ and $F(1, 45) = 10.81$, $MSE = .02$, $p < .01$, but this decrease was larger for the CS+ than for the CS-. In terms of conditioned fear this means that a reduction in fear was observed throughout extinction for both CSs with a particularly strong decrease in fear for the CS+. In addition, the analysis yielded significant main effects of moment, $F(1, 45) = 211.70$, $MSE = .01$, $p < .001$, and CS-type, $F(1, 45) = 80.34$, $MSE = .02$, $p < .001$. Overall, less suppression (i.e., less conditioned fear) was observed at the end of extinction in comparison with the end of acquisition and the CS+ elicited more suppression/fear than the CS-. At the end of extinction there was no significant difference between the mean suppression ratios for the CS+ ($M = .48$) and CS- ($M = .47$), $F < 1$. Hence, the CS+ and CS- no longer differed with respect to the amount of conditioned fear they elicited. None of the effects interacted with condition or treatment.

We can conclude that extinction has clearly taken place. In line with our predictions, a strong decrease in conditioned fear for the CS+ was observed throughout extinction. At the end of extinction both CSs elicited similar and low amounts of fear. The preconditions for investigating reinstatement (i.e., successful differential acquisition and extinction in all groups) were thus met.



a



b

Figure 10. Mean suppression ratios for the last (20) extinction session and the test session (T) for the FE-reinstatement group (left hand side of a), the FE-control group (right hand side of a), the CC-reinstatement group (left hand side of b) and the CC-control group (right hand side of b). Lower values represent stronger conditioned responding. Error bars represent standard errors of means.

Test of reinstatement

The reinstatement manipulation was expected to lead to a return of conditioned fear for the CS+ but not for the CS- in the reinstatement groups, while the control groups were not expected to show any changes in comparison with responding during the extinction phase. Importantly, we expected to observe less return of fear for the CS+ in the CC-reinstatement group as compared to the

FE-reinstatement group. In terms of suppression ratios this means that we expected a decrease in the suppression ratio for the CS+ in the reinstatement groups, with a stronger decrease for the FE-reinstatement group as compared to the CC-reinstatement group.

To test for reinstatement, the mean suppression ratio of the last extinction session was compared with that of the test session. Figure 10 shows the mean suppression ratios of the last extinction session and the test session per group. A strong return of conditioned responding for the CS+ was observed in the reinstatement groups. Contrary to what we expected, increased suppression (i.e., fear) was also observed for the control groups. Moreover, and also in contrast to our expectations, a similarly strong return of fear was observed for the CC groups as compared to the FE groups. A 2 (Condition) x 2 (Treatment) x 2 (CS-type) x 2 (Moment) ANOVA confirmed this description of the data. The crucial test for differential reinstatement, the Moment x CS-type x Condition interaction failed to approach significance, $F(1, 45) = 2.03$, $MSE = .02$, $p = .16$. Importantly, the Condition x Moment interaction was significant, $F(1, 45) = 9.30$, $MSE = .03$, $p < .01$, indicating that the increase in fear was stronger in the reinstatement groups than in the control groups. The ANOVA revealed main effects of all three variables. Overall, the CS+ elicited more suppression/fear than the CS-, $F(1, 45) = 20.38$, $MSE = .02$, $p < .001$. More suppression/fear was observed during the test session in comparison with the end of extinction, $F(1, 45) = 60.56$, $MSE = .03$, $p < .001$, and more suppression/fear occurred in the reinstatement groups as compared to the control groups, $F(1, 45) = 12.56$, $MSE = .02$, $p < .001$. Further, a significant Moment x CS-type interaction was observed, $F(1, 45) = 20.36$, $MSE = .02$, $p < .001$, that indicated a stronger increase in suppression from the end of extinction to test for the CS+ as compared to the CS-. In other words, a stronger return of conditioned fear was observed for the CS+ than for the CS-. The fact that this interaction was not only significant in the reinstatement groups, $F(1, 45) = 17.99$, $MSE = .02$, $p < .001$, but also in the control groups, $F(1, 45) = 4.67$, $MSE = .02$, $p < .05$, signifies that differential spontaneous recovery of conditioned fear occurred in those last groups. Next, we analyzed the data separately for the CS+ and CS-. A significant Moment x Condition interaction was observed for the CS+, $F(1, 45) =$

10.47, *MSE* = .02, $p < .01$, with a stronger increase in suppression/fear for the CS+ in the reinstatement groups than in the control groups. This interaction was not significant for the CS-, $F(1, 45) = 2.00$, *MSE* = .02, $p = .16$. Hence, even though a differential ROF was observed both in the reinstatement groups and the control groups, the unsignaled USs in the reinstatement groups resulted in a stronger ROF for the CS+.

Next, we looked at whether this reinstatement effect was modulated by our CC manipulation. The Treatment x Condition x Moment interaction was, however, not significant for the CS+, $F(1, 45) = 1.14$, *MSE* = .02, $p = .29$. Hence, contrary to what we predicted, a similar amount of return of fear for the CS+ after reinstatement was observed in the CC-reinstatement group and the FE-reinstatement group. As we hypothesized that CC might also be effective in reducing return of fear due to spontaneous recovery, we compared the control groups to see if less spontaneous recovery for the CS+ alone or for the CS+ as compared to the CS- could be observed in the CC-control group as compared to the FE-control group. This was, however, not the case, both $F < 1$. Thus, in contrast to what we hypothesized, no effect was found of the CC treatment on ROF due to reinstatement or spontaneous recovery.

Conclusion

In the present study we wanted to investigate the impact on ROF of applying a CC procedure to diminish the negative valence of the CS+ after extinction. In a differential conditioned suppression paradigm in mice, successful differential acquisition and extinction of conditioned suppression were observed. Low amounts of fear were observed for the CS+ and CS- at the beginning of acquisition. Throughout the acquisition phase the mice gradually acquired the differential contingencies between the CSs and the US and showed an increase in conditioned fear for the CS+ but not for the CS-. The repeated presentation of the CSs in absence of the US led to the extinction of this differential fear. The preconditions for investigating reinstatement were thus met.

We failed, however, to replicate the differential reinstatement effect found by Dirikx et al. (2007). This was mainly due to the fact that a differential ROF (with a stronger return of conditioned responding for the CS+ as compared to the CS-) was not only observed in the reinstatement groups, but also in the control groups that had received no unsignaled USs. This phenomenon of a return of extinguished conditioned responding merely through the passage of time is known in the literature as 'spontaneous recovery' (for a review see Rescorla, 2004). In general, stronger spontaneous recovery effects are observed when the time delay between extinction treatment and test becomes longer (e.g., Quirk, 2002). The spontaneous recovery effect in the control groups of the current study is then probably due to the 12 sessions nose poke retraining that took place (and hence formed a delay period) in between the CC/FE phase and the reinstatement phase and during which no CS tones were presented. Future research could add a short extinction reminder phase (consisting of a few extinction sessions) before the test phase to prevent the occurrence of spontaneous recovery in the control groups. When the data were analyzed separately for the CS+ and CS-, a stronger ROF for the CS+ was observed in the reinstatement groups than in the control groups. Hence, we can say that our reinstatement manipulation was effective and resulted in additional ROF for the CS+ in the reinstatement groups as compared to the control groups.

In contrast to what we predicted, no beneficial effect was found of the CC treatment on ROF due to reinstatement. An equally strong return of conditioned responding after the reinstatement manipulation was observed both in the FE-reinstatement group and in the CC-reinstatement group. As we hypothesized that CC might also be effective in reducing ROF due to spontaneous recovery, we compared the control groups to see if less spontaneous recovery could be observed in the CC-control group as compared to the FE-control group. This was not the case. Hence, we can conclude that our CC treatment was ineffective in reducing ROF due to reinstatement or spontaneous recovery.

Taken together, the data of the present study support previous studies that found that ROF can still be observed after applying a CC procedure (Bouton & Peck, 1992; Brooks et al., 1995, Peck & Bouton, 1990). In addition, the present data suggest that CC is also ineffective in *reducing* ROF by changing the negative valence of the CS+ after extinction. One possible interpretation of this finding is that the hypothesis that the extinction-resistant negative valence of the CS+ forms a source for ROF should be rejected. If we translate this to the clinical practice, our findings then suggest that combining standard exposure techniques with procedures which are specifically designed to reduce the negative valence of the fear-eliciting stimulus, not necessarily results in a lower relapse rate. An alternative explanation of our null finding, however, is that the valence of the CS+ was perhaps not (sufficiently) altered by our CC procedure. Because the valence of the CSs was not measured during the experiment, this alternative explanation cannot be excluded. The CC phase of the present experiment was longer than in other animal studies (e.g., Bouton & Peck, 1992; Brooks et al., 1995; Delprato & Jackson, 1973; Peck & Bouton, 1990). Nevertheless, the 17 sessions of CC might still have been insufficient to alter the negative valence of the CS+. Although there is evidence that very short contingent presentations are already sufficient to change the affective valence of neutral stimuli (e.g., Baeyens, Eelen, Crombez, & Van den Bergh, 1992) or to neutralize the valence of previously conditioned stimuli (e.g., Baeyens et al., 1989), it may well be that a longer CC procedure or the contingent presentation of more intense or several positively valenced stimuli are necessary in case of stimuli with a very strong negative valence. In order to learn more about the role of stimulus valence in fear conditioning in animals, it will be important for further research to develop a valence/preference measure that can be administered during the experiment. Finding a good index for evaluative learning (certainly in the case of tone-CSs) is, however, not a simple task as it is often difficult to disentangle evaluative learning from expectancy learning.

To conclude, the present results showed that ROF was not reduced by applying a CC procedure to diminish the negative valence of the CS+ after extinction. However, as we cannot definitely conclude that our CC procedure succeeded in changing the valence of the CS+, more research is needed to verify this conclusion.

References

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996) No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation, 27*, 200-241.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning: Acquisition trials, presentation schedule, evaluative style, and contingency awareness. *Behaviour Research and Therapy, 30,* 133-142.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective evaluative value - Conservative but not unchangeable. *Behaviour Research and Therapy, 27*, 279-287.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*, 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*, 485-494.

Bouton, M. E., & Peck, C. A. (1992). Spontaneous-recovery in cross-motivational transfer (counterconditioning). *Animal Learning & Behavior, 20*, 313-321.

Brooks, D. C., Hale, B., Nelson, J. B., & Bouton, M. E. (1995). Reinstatement after counterconditioning. *Animal Learning & Behavior, 23*, 383-390.

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin, 25,* 215-224.

Craske, M. G., Hermans, D., & Vansteenwegen, D. (2006). *Fear and learning: Basic science to clinical application.* Washington, DC: American Psychological Association.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: some current controversies and possible ways forward. *Cognition & Emotion, 19*, 161-174.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127,* 853-869.

de Jong, P. J., Vorage, I., & van den Hout, M. A. (2000). Counterconditioning in the treatment of spider phobia: Effects on disgust, fear, and valence. *Behaviour Research and Therapy, 11*, 1055-1069.

Delprato, D. J., & Jackson, D. E. (1973). Counterconditioning and exposure-only in treatment of specific (conditioned suppression) and generalized fear in rats. *Behaviour Research and Therapy, 11*, 453-461.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self-esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345-355.

Dirikx, T., Beckers, T., Muyls, C., Eelen, P., Vansteenwegen, D., Hermans, D., & D'Hooge, R. (2007) Differential acquisition, extinction and reinstatement of conditioned suppression in mice. *The Quarterly Journal of Experimental Psychology, 60*, 1313-1320.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*, 549-554.

Eifert, G. H., Craill, L., Carey, E., & O'Connor, C. (1988). Affect modification through evaluative conditioning with music. *Behaviour Research and Therapy, 26,* 321-330.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229-238.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Hamm, A. O., Vaitl, D., & Lang, P. J. (1989). Fear conditioning, meaning, and belongingness - a selective association analysis. *Journal of Abnormal Psychology, 98,* 395-406.

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In S. P. Shohov (Ed.), *Advances in Psychology Research, Vol. 12* (pp.17-41). New York: Nova Science.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*, 533-551.

Huijding, J., & de Jong, P. J. (2005). *Changes of automatic and self-reported spider phobia-related affective associations following treatment and the prediction of symptom return*. Unpublished doctoral dissertation, Rijksuniversiteit Groningen.

Kessler, R. C., Koretz, D., Merikangas, K. R., & Wang, P. S. (2004). The epidemiology of adult mental disorders. In B. L. Levin, J. Petrilia & K. D. Hennessy (Eds.), *Mental health services: A public health perspective, Second Edition.* New York: Oxford University Press.

Korrelboom, C. W., van der Gaag, M., Hendriks, V. M., Huijbrechts, I. P. A. M., & Berretty, E. W. (2008). Treating obsessions with Competitive Memory Training: A pilot study. *The Behavior Therapist, 31*, 29-36.

Lang, P. J. (1995). The emotion probe - studies of motivation and attention. *American Psychologist, 50,* 372-385.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 431-440.

Marks, I. M. (1987). *Fears, phobias, and rituals: Panic, anxiety and their disorders.* Oxford: Oxford University Press.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*, 421-433.

Öst, L. G. (1989). One-session treatment for specific phobias. *Behaviour Research and Therapy, 27,* 1-7.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation, 21,* 1-31.

Quirk, G. J. (2002). Memory for extinction of conditioned fear is long-lasting and persists following spontaneous recovery. *Learning and Memory, 9,* 402-407.

Rachman, S. (1989). The return of fear: Review and prospect. *Clinical Psychology Review, 9,* 147-168.

Rescorla, R. A. (2004). Spontaneous recovery. *Learning and Memory, 11,* 501-509.

Rydell, R. J., & McConnell, A. R. (2006). Understanding implicit and explicit attitude change: A systems of reasoning analysis. *Journal of Personality and Social Psychology, 91,* 995-1008.

Rydell, R. J., McConnell, A. R., Strain, L. M., Claypool, H. M., & Hugenberg, K. (2007). Implicit and explicit attitudes respond differently to increasing amounts of counterattitudinal information. *European Journal of Social Psychology, 37*, 867-878.

Scalera, G. (2000). Taste preference and acceptance in thirsty and dehydrated rats. *Physiology and Behavior, 71*, 457-468.

van Balkom, A. J. L. M., Bakker, A., Spinhoven, P., Blaauw, B. M. J. W., Smeenk, S., & Ruesink, B. (1997). A meta-analysis of the treatment of panic disorder with or without agoraphobia: A comparison of psychopharmacological, cognitive-behavioral, and combination treatments. *Journal of Nervous and Mental Disease, 185,* 510-517.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 71-79.

3 Assessing valence indirectly and online

As is clear from the first part of this dissertation, the main focus of my research was on evaluative learning, and more specifically on how conditioned (dis)likes are (formed and) changed. Obtaining a good measurement of people's evaluations is crucial if one wants to study evaluative learning processes. Against this background, we focused in a second line of research on the assessment of valence and developed an indirect reaction time (RT) measure that can be used to assess participants' evaluations online, during evaluative learning. Three experimental studies will be discussed in the following chapters.

The first chapter provides an overview of the measurements that are commonly used in experimental research to assess participants' evaluations. The most popular valence measures in evaluative conditioning (EC) studies are self-report ratings and indirect RT tasks. In EC studies, these measures are typically administered in a session that precedes and/or follows the EC phase (i.e., as pre-post measures). In Chapter 1, we discuss the disadvantages of such pre-test post-test designs and argue that it might be preferable (or at least interesting) to assess valence *during* conditioning, rather than only before and after. As the currently existing behavioral valence measures do not lend themselves well to be integrated in an ongoing conditioning procedure, we focused in this second line of research on the development of an indirect RT task that *can* be administered online. In our search to develop this measure, we were inspired by the work of Dawson, Beers, Schell, and Kelly (1982), who measured the operation of cognitive processes in a conditioning procedure by asking participants to respond to (neutral) tone probes that were presented during the presentation of the CSs. We developed an affective variant of this measure by replacing the neutral tone probes with affect-laden visual probes. The in this way obtained task can be conceptualized as an online variant of the affective priming task.

In Chapter 2 a pilot study is presented in which we tested whether our measure is able to pick up the valence of normatively selected positive and negative pictures. Results supported the validity of our task.

In Chapter 3 we discuss a study in which we examined whether this RT task is capable of assessing participants' evaluations online, during evaluative learning. We therefore applied the task in a differential fear acquisition and subsequent extinction procedure with an electrocutaneous stimulus as US. Results indicated that our task succeeded in tracking the expected changes in valence of the CSs.

As indicated above, the RT task we developed is based on the secondary probe RT task of Dawson et al. (1982), which has been used as an online measure of expectancy/fear in several conditioning studies. In this second experiment, we expected our adapted version of this task to not only track evaluative learning, but to still measure expectancy learning too. We failed, however, to observe any effect of expectancy learning in our task.

Chapter 4 presents a third and final study, in which we replicated the procedure of Experiment 2, but made same crucial changes that were aimed at making the expectancy component in the experiment more salient. Results indicated that our task again succeeded in picking up evaluative learning but not expectancy learning. Potential reasons for the latter finding are discussed in the chapter.

Each chapter includes a discussion section. The results of this second line of research will, however, also be recapitulated and discussed in the General Discussion of this dissertation (Part 4).

Assessing valence

An overview of the most commonly used valence measures in evaluation research

For many years, psychologists have tried to measure people's evaluations in an attempt to understand, control or predict human behavior. Most often, they have done so using self-report questionnaires in which respondents are asked to express their attitude towards a certain object, activity, person or group by marking a rating scale or selecting one of several response alternatives (e.g., 'indicate on a scale ranging from -100 (very negative) to +100 (very positive) how you feel towards smoking').[34] In the literature, this type of measurements is referred to as *direct* attitude measures because they require the respondent to self-assess the to-be-measured attitude (De Houwer & Moors, in press). Direct attitude measures have the advantage of being fairly easy to construct and administer. Disadvantages are the fact that such measures might be strongly biased by interpretation errors (e.g., respondents can misinterpret the questions or response alternatives, see Schwartz, 1999), social desirability and self-presentational strategies (e.g., Dovidio & Fazio, 1992). Direct measures are further based on the assumption that respondents have introspective access to their evaluations and are aware of what they like and dislike. Not all attitude researchers are convinced that this is the case.

To overcome the problems that are associated with direct measures, *indirect* measures were developed. In the latter type of measurements, respondents are not directly asked to report their attitude towards a certain stimulus. Instead, the respondent's evaluation is inferred from his/her behavior or responses. Several behavioral, physiological and neurological measurement procedures were developed in the last decades to assess people's evaluations without relying on direct self-report (see further). The main reason for why many researchers are attracted to indirect measures, is that these measures are often assumed to be *implicit*. The concepts 'implicit' and 'implicit measure' have been used in several ways in the literature. According to De Houwer (2006), contemporary attitude researchers most commonly use the term 'implicit attitude measure' to refer to a measure that provides an index of a certain attitude even

[34] Various types of self-report scales have been used: semantic differentials (Osgood, Suci, & Tannenbaum, 1957), Likert scales (Likert, 1932), VAS scales, etc.

though respondents (a) are not aware of the fact that the attitude is being measured (e.g., Brunel, Tietje, & Greenwald, 2004), (b) do not have conscious access to the attitude (e.g., Asendorpf, Banse, & Mücke, 2002), or (c) have no control over the measurement outcome (e.g., Fazio & Olson, 2003).[35] It is clear that indirect measures are more likely to possess (one of) these functional properties than direct measures. Asking respondents to self-assess the to-be-measured attitude renders it likely that participants will become aware of what is being measured and/or are able to control the measurement outcome. Note, however, that the terms direct-indirect and explicit-implicit are not synonyms.[36] Direct measures are not by definition explicit measures. De Houwer (2006) gives the example that one can ask participants to express their liking of a certain object as quickly as possible and/or while performing a demanding secondary task. In such cases, participants might have little control over the expressed evaluation (e.g., Wilson, Lindsey, & Schooler, 2000). Conversely, indirect measures are not by definition implicit measures. Whether a certain indirect measure possesses one (or more) of the functional properties that are seen as characteristic for an implicit measure should be examined empirically (De Houwer, 2006, 2009; De Houwer & Moors, 2007).

In the following section, we look at some indirect measures that are commonly used in evaluation research. A classification is made according to the type of responses the indirect measure relies on: (a) behavioral responses (e.g., spoken or written answers, key press responses), (b) physiological responses (e.g., startle reflex), or (c) neurological responses (e.g., brain activity as registered by EEG).

[35] In some recent articles, De Houwer (2006, 2009; De Houwer & Moors, 2007) pointed out that even though the terms ‘implicit’ and ‘implicit measure’ are commonly used, few authors specify what they mean exactly by it [e.g., one (or more) of the here mentioned functional properties or even other properties]. Because the term ‘implicit’ has been associated with a variety of functional properties in the literature (properties that do not necessarily co-occur!), De Houwer and colleagues suggest that, in order to avoid misunderstandings, authors should always specify to which functional property (or properties) they refer when they use the term implicit.

[36] In the present chapter we follow the conceptual framework and definitions of De Houwer (2006). More information on the distinction between the qualifications ‘direct/indirect’ and ‘explicit/implicit’ can be found in his paper.

Behavioral responses

The currently most widely used indirect attitude measures rely on response time measurement. The two most popular response time measures are the Affective Priming Task (APT) developed by Fazio, Sanbonmatsu, Powell, and Kardes (1986) and Greenwald, McGhee, and Schwartz's (1998) Implicit Association Test (IAT).

Historically, the APT can be regarded as the first and theoretically most influential of these types of measures. In an APT, two stimuli are presented consecutively and participants are asked to evaluate the second stimulus, that is, to determine whether this second stimulus refers to something good or something bad (e.g., by pressing one button for good words and another one for bad words). Results consistently show that less time is needed to evaluate the second stimulus - also called the target - when this stimulus has the same valence as the first stimulus - also called the prime - (e.g., puppy - happy) than when both stimuli have a different value (e.g., dead - happy) (e.g., Bargh, Chaiken, Govender, & Pratto, 1992; Fazio et al., 1986; Hermans, De Houwer, & Eelen, 1994). The APT can and has been used to measure participants' evaluations. If the presentation of the prime leads to faster processing of positive than negative targets, one can infer that the participant has a positive attitude towards the prime stimulus. If a prime facilitates the processing of negative compared to positive targets, this suggests that the prime has a negative connotation for the participant. Priming effects have typically been explained by a 'spreading of activation' mechanism (Collins & Loftus, 1975). The idea behind this account is that exposure to a certain stimulus (e.g., the prime) activates semantically related concepts in memory (e.g., concepts that share the same valence like the target stimulus), thereby reducing the time needed to identify these concepts. More recent evidence, however, suggests that response competition/facilitation might play an important role in priming effects (see Klauer & Musch, 2003; Klauer, Musch, & Eder, 2005 for reviews). According to this view, the prime activates a response on the basis of its valence which 'readies' the individual to respond accordingly. If the subsequently presented target is congruent with the prime, responding is facilitated because the response pathway already has received some initiation. On the other hand, if the target is evaluatively incongruent, then the response suggested by the evaluation associated with the prime must be inhibited in order to respond accurately to the target. Thus, the evaluation activated by the prime and that activated by the target either may complement one another and, hence, facilitate responding, or they may conflict with one another and, hence interfere with responding. For a more detailed review on the possible underlying processes of priming effects, we refer to Wittenbrink (2007) and De Houwer, Teige-Mocigemba, Spruyt, and Moors (2009).

The second most popular response time measure is the IAT. The IAT is a sorting task predicated on the assumption that it is easier to make the same response to two things when they are related than whey they are unrelated. The IAT involves presenting the participants with two attribute concepts such as 'positive' and 'negative', and two target concepts such as 'male' and 'female'. Target stimuli pertaining to these four concepts (e.g., 'love', 'dead', 'Mary', and 'John') are then presented on a computer screen, and participants' task is to respond to these stimuli by selecting the relevant response key. In one task, 'positive' and 'female' are grouped together and share a single response key, as do 'negative' and 'male'. In a second task, the response assignment for 'male' and 'female' is reversed while the assignment for 'positive' and 'negative' remains the same. Thus now 'positive' and 'male' share a response key, as do 'negative' and 'female'. The critical measure (the IAT effect) is the difference in response times between these two tasks. Faster responses in the first as compared to the second task are assumed to indicate a more positive evaluation of females as compared to males, while the reverse pattern suggests that the respondent evaluates males more positively than females. Even though the IAT has become widely accepted and used as an indirect attitude measure, relatively little is know about its underlying processes. Several theoretical accounts have been proposed to explain IAT effects (see De Houwer et al., 2009 for a review). In many of these accounts the degree to which the target and attribute categories share similar features plays a key role. De Houwer (2003b), for instance, suggests that IAT effects may draw on a stimulus-response compatibility mechanism whereby target and attribute information elicit synergistic response tendencies when similar categories are mapped on the same response key. In contrast, when dissimilar categories are mapped on the same response key, the response tendencies for target and attribute categories are antagonistic to each other, resulting in slower response times. In a similar vein, the task-switching account by Mierke and Klauer (2001, 2003) states that the pairing of dissimilar categories requires participants to switch between target and attribute discrimination. In contrast, participants only need to consider attribute-related information if associated categories are paired with each other. Research on task switching has shown that performance deteriorates as the result of switching between tasks (e.g., Meiran, Chorev, & Sapir, 2000), which could then explain why participants are slower on IAT trials with an ‘incompatible’ response key assignment. An overview of other mechanisms that have been postulated to underlie IAT effects can be found in De Houwer et al. (2009).

In the last two decades, a number of other reaction time (RT) tasks have been introduced that also provide potential ways to measure attitudes indirectly. These tasks include the (extrinsic) affective Simon task (e.g., De Houwer, 2003a; De Houwer & Eelen, 1998), the emotional Stroop task (e.g., Pratto & John, 1991), the single-category IAT (Karpinski & Steinman, 2006), the Go/No-go Association Task (GNAT; Nosek & Banaji, 2001), the single association test (Blanton, Jaccard, Gonzales, & Christie, 2006), the stimulus response compatibility task (Mogg, Bradley, Field, & De Houwer, 2003), the sorting paired features task (Bar-Anan, Nosek, & Vianello, 2009) and the brief IAT (Sriram & Greenwald, 2009). A detailed review of each of these tasks is beyond the scope of this chapter, but more information can be found in the cited articles.

Most but not all indirect behavioral measures focus on participants’ speed of responding to make inferences about their evaluations. For some of the above mentioned tasks, accuracy scores (instead of response latencies) have sometimes been used as an index of participants’ evaluations (e.g., the GNAT, Teachman, 2006). Another example is the Affect Misattribution Procedure of Payne, Cheng, Govorun, and Stewart (2005), in which inferences about people’s attitudes are made based on how they judge ambiguous objects after being exposed to the attitude object of real interest. The underlying idea is that a person’s reaction to a target object (e.g., a picture of a cute baby) can be misattributed to the subsequently presented ambiguous object (e.g., a Chinese ideograph). Note that even though an individual's evaluation is inferred from self-reported ratings in this task, it can be considered an indirect measure because participants are not directly asked for their evaluation of the target object (i.e., they provide direct ratings for the ambiguous stimuli).

Physiological responses

Because of their involuntary and hard to control nature, physiological correlates have captured the interest of many attitude researchers who doubted respondents' explicit self-reports.

Autonomic measures. Early uses of physiological measures drew on the observation that strong affective reactions to a stimulus are associated with increased activation of the sympathetic nervous system (e.g., Rankin & Campbell, 1955). Increased activation of this system is accompanied by elevated sweat gland activity, which can be measured by assessing the skin's resistance to low level electric currents, a procedure known as electrodermal measurement. Electrodermal responses, however, do not reflect the direction (favorable or unfavorable) of the evaluative response, which limits their usefulness as indirect valence measure (e.g., Cacioppo & Tassinary, 1990). Electrodermal activity can therefore better be conceptualized as an index of arousal or affective intensity.

Within the evaluation literature, there is a great deal of research examining the potential for indices of cardiovascular activity, such as heart rate, to discriminate between positive and negative evaluative responses. Some early studies found heart rate changes to vary as a function of affective valence with unpleasant pictures prompting slower heart rates than pleasant ones (e.g., Greenwald, Cook, & Lang, 1989; Lang, Greenwald, Bradley, & Hamm, 1993). More recent studies looked at the time course of heart rate during picture viewing and identified a triphasic pattern with an initial deceleration followed by an accelerative component and subsequently a secondary deceleration. This triphasic response was modulated by affective valence with a more pronounced initial deceleration for unpleasant pictures and a larger accelerative peak for positive pictures (e.g., Bradley, Codispoti, Cuthbert, & Lang, 2001). Other studies, however, failed to replicate this pattern of results (e.g., Rottenberg, Kasch, Gross, & Gotlib, 2002; Levenson & Ekman, 2002) which led some authors to conclude that heart rate is not a reliable indicator of stimulus valence (e.g., Ito & Cacioppo, 2007; Schimmack & Crites, 2005).

Skeletal nervous system measures. A more promising approach involves the measurement of facial electromyographic (EMG) activity. People often show overt facial expressions (like smiling or frowning) to attitude objects that elicit strong evaluative reactions. These expressions may, however, be intentionally concealed and many evaluative reactions are too subtle to evoke such overt expressive behaviors. Nevertheless, research indicates that even subtle evaluative reactions are associated with a low-level activation of certain specific facial muscles that can be detected by EMG (e.g., Tassinary & Cacioppo, 1992). For this purpose, small electrodes are placed on the surface of the skin over the facial muscles of interest. Using EMG, several studies have found that the processing of unpleasant events is associated with greater activity over the corrugator (frown) muscle while processing pleasant events prompts greater activity over the zygomatic (smile) muscle (e.g., Lang et al., 1993; Schwartz, Brown, & Ahern, 1980; Tassinary, Cacioppo, & Geen, 1989). Based on these findings, some authors have used facial EMG activity as an indirect index of stimulus valence (e.g., Wilbarger, McIntosh, & Winkielman, 2009).

Another regularly used physiological measure relates to the startle response, which is an involuntary defensive reflex that occurs in response to a sudden intense stimulus (e.g., a loud noise) and involves multiple motor actions, including tensing of the neck and back muscles and blinking of the eyes. The most robust component of the behavioral cascade that constitutes the startle reflex is the eye blink. In experimental contexts, this startle eye blink reflex can be evoked by presenting participants with a short, intense blast of noise (referred to as the startle probe) and is typically measured by a pair of electrodes placed below the eye, over the orbicularis oculi muscle. Multiple studies suggest that the magnitude and latency of the startle eye blink reflex varies as a function of stimulus valence (for reviews, see Lang, Bradley, & Cuthbert, 1990 and Grillon & Baas, 2003). Vrana, Spence, and Lang (1988), for instance, have demonstrated that larger and faster blinks are elicited by startle probes that occur during exposure to unpleasant stimuli, whereas smaller, slower blinks are observed in response to probes presented during pleasant stimuli. Lang and colleagues have proposed a motivational priming account for these affective modulation effects. In this view, an individual's current affective state augments motivationally congruent reflexive responses but inhibits incongruent ones (Lang, 1995; Lang et al., 1990; Lang, Bradley, & Cuthbert, 1992). The startle response is considered a defensive reflex, resulting in facilitation (inhibition) of the reflex when presented in the context of a negative (positive) affective foreground stimulus. The startle eye blink reflex provides a promising indirect measure of stimulus valence. It should be noted, however, that some studies (e.g., Cuthbert, Bradley, & Lang, 1996) found affective modulation of this reflex to occur only for highly arousing stimuli. Therefore, the startle eye blink reflex might be less suitable to differentiate between stimuli of moderate affective intensity.

Neurological responses

Several studies have explored whether brain activity can be used to assess a person's evaluations. One approach involves the measurement of brain activity through electroencephalography (EEG), which entails the recording of small electrical signals along the scalp. This procedure, however, does not lend itself to a direct assessment of positive or negative responses. Instead, it capitalizes on the observation that unexpected stimuli evoke brain wave activity that differs from the activity evoked by expected stimuli. Cacioppo, Crites, and Gardner (1996) found that the amplitude of a particular type of wave form - the Late Positive Potential or LPP - varies as a function of evaluative (in)consistency. A positive (negative) target stimulus presented within a series of negative (positive) pictures, for instance, elicits a larger LPP than does the same target presented among other positive (negative) stimuli. Hence, one can detect whether a target object is evaluated positively or negatively by embedding its presentation in a long series of other objects with a known evaluation. The amplitude of the LPP evoked by the target object will then indicate if its evaluation is consistent or inconsistent with the evaluation of the context objects (e.g., Cacioppo et al., 1996; Cacioppo, Crites, Berntson, & Coles, 1993; Ito &

Cacioppo, 2000). The amplitude of the LPP wave is, moreover, also sensitive to the degree of evaluative inconsistency. A stimulus that is extremely incongruent with the preceding stimuli (e.g., a very positive item shown among a majority of very negative items) will elicit a larger LPP than a moderately evaluatively incongruent stimulus (e.g., a moderately positive item shown among a majority of very negative items) (e.g., Cacioppo et al., 1996; Cacioppo, Crites, Gardner, & Berntson, 1994).

In the last 10 years, there has been a rapid expansion in cognitive neuroscience research using functional magnetic resonance imaging (fMRI) to study the brain correlates of several psychological processes, including evaluation. Through fMRI, changes in blood flow related to neural activity can be measured. fMRI has been applied to the study of attitudes only recently, with most studies focusing on the role of the amygdala: a small almond-shaped structure in the medial temporal lobe at the tip of the hippocampus. Hitherto, it is not yet clear how valence and amygdala activation are related (Ito & Cacioppo, 2007). Increases in amygdala activation are most consistently associated with negative stimuli (e.g., Isenberg et al., 1999; Morris et al., 1996; Paradiso et al., 1999), but greater activation is sometimes also observed in response to positive as compared to neutral stimuli (e.g., Garavan, Pendergrass, Ross, Stein, & Risinger, 2001; Hamann, Ely, Hoffman, & Kilts, 2002; Hamann & Mao, 2002; Liberzon, Phan, Decker, & Taylor, 2003). Whalen (1998) has suggested that amygdala activation in response to negative stimuli might occur because the amygdala is important for modulating vigilance. From this perspective, amygdala activation may be seen most consistently in response to negative stimuli, but should also occur in response to any other stimulus signaling biological relevance. Another explanation is given by Cunningham, Raye, and Johnson (2004), who noted that the valence and intensity dimensions of stimuli used in fMRI studies have not always been separated consistently. The tendency for negative stimuli to often be more intense than positive stimuli (cf. Ito & Cacioppo, 2000) raises the possibility that the increased amygdala activation to negative stimuli relates to their higher intensity. Three fMRI studies that disentangled the effects of valence and arousal indeed found that only stimulus intensity predicted amygdala activation (Anderson & Sobel, 2003; Cunningham et al., 2004; Small et al., 2003). Hence, these data suggest that increased amygdala activation cannot be taken as unequivocal evidence that a stimulus is evaluated more negatively.

Several other brain areas have been associated with the processing of valence (e.g., the ventral striatum and the right orbital frontal cortex, see Ito & Cacioppo, 2007 for a review). More research is needed, however, to explore whether activity in any of these regions can function as a reliable indirect index of stimulus valence.

Substantial disadvantages of both physiological and neurological measures are the fact that their implementation requires high expertise and sophisticated technology and that the analysis of the data resulting from these measures is often fairly complex. Behavioral measurements like RT tasks, on the other hand, require only the technology of personal computers and provide data that are easy to collect and analyze. This probably explains the widespread use of the latter type of measures in evaluation research. In the following section, we describe which valence measures are generally applied in EC research and explain why we felt the need to develop a new RT measure.

Valence measures in EC research: the need for an online indirect reaction time task

Almost all of the above reviewed valence indices have been applied in EC research (e.g., facial EMG, Hermann, Ziegler, Birbaumer, & Flor, 2000; modulation of the startle reflex, Purkis & Lipp, 2001; fMRI responses, Klucken et al., 2009). Nevertheless, two types of measurements have prevailed in the EC literature: self-report ratings and indirect RT tasks.

Traditionally, EC studies have relied on direct measures of liking. In the vast majority of EC studies, participants were asked to self-assess their liking of the CSs and USs, for instance, by selecting a number on a Likert scale or by sorting the stimuli into separate piles for liked, neutral and disliked stimuli (e.g., Levey & Martin, 1975; Baeyens, Eelen, Crombez, & Van den Bergh, 1992; for reviews, see De Houwer, Thomas, & Baeyens, 2001; De Houwer, Baeyens, & Field, 2005; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, in press). In the last two decades, indirect RT tasks like the APT and IAT were introduced to, and rapidly gained widespread use in, the field of EC research. The fast breakthrough of this type of measurements probably relates to the fact that they are assumed to be more implicit than self-report ratings (also see above) and are easy to apply. Developed in the context of social psychology research, indirect RT measures were typically used to assess attitudes that are considered to be highly stable and long-lasting (e.g., political attitudes, Fazio & Williams, 1986; racial attitudes, Greenwald et al., 1998; self-esteem, Greenwald & Farnham, 2000). Several EC studies, however, found these measures to also be capable of assessing (changes in) newly acquired attitudes (e.g., De Houwer, Hermans, & Eelen, 1998; Olson & Fazio, 2001).

To study (changes in) evaluative learning, most EC studies employ a pre-test post-test design in which ratings and/or an indirect RT task are administered in a session that precedes and/or follows the conditioning phase. An example can be found in the studies that were presented in Part 2. In all of these studies, shifts in valence were examined by comparing rating (or priming) data collected before and after conditioning. Several authors have pointed out, however, that pre-test post-test studies can have a number of disadvantages. This type of studies, for instance, provides little insight into the *course* of evaluative learning as only the beginning and end result of the learning process is evaluated (Blechert, Michael, Williams, Purkis, & Wilhelm, 2008). Pre/post measures remain silent about the slope and shape of participants' learning curve. A second disadvantage relates to the repeated administration of indirect RT tasks like the APT and IAT. In several studies smaller APT and IAT effects have been observed when participants performed these tasks more than once (e.g., Experiment 3 in Part 2; Greenwald, Nosek, & Banaji, 2003). It is not yet clear why prior experience with the APT and IAT is associated with smaller effects. Floor-effects (e.g., caused by the fact that participants become better and faster at the task) might possibly contribute to this phenomenon. Another explanation might be found in a recent study by McDaniel, Beier, Perkens, Goggin, and Frankel (2009), who found that prior experience with the IAT can increase participants' control over their responses. Regardless of what exactly causes the reduced effects, this finding poses a serious problem for studies that want to examine changes in liking by repeatedly administering an indirect RT task. A third and even more problematic aspect of pre-test post-test designs was identified by Lipp, Oughton, and Lelievre (2003). These authors pointed out that in pre-test post-test studies, the transition from the learning phase to the test phase to some extent always entails a change in context. The test phases often take place under conditions that differ from those of the conditioning phase (e.g., change in room or test material) and even the mere transition of time between the conditioning phase and the measurement could suffice to constitute a context change (Bouton, 2004). As learning effects can be sensitive to context changes (for reviews see Blair, 2002; Ferguson & Bargh, 2007; Schwartz, 2008), Lipp et al. (2003) suggest that the results of pre-test post-test studies might be biased by context effects. According to the authors, an example might be found in studies on the extinction insensitivity of evaluative learning. The experiments that report extinction-resistant evaluative learning all employed a pre-test post-test design (e.g., Diaz, Ruiz, & Baeyens, 2005; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006). In Lipp et al.'s (2003) view the post-tests in these studies might reflect context-induced *renewal*[37] of previously extinguished evaluative learning rather than extinction-resistant evaluative learning. A final argument against the use of pre-test post-test designs comes from the causal learning literature. Collins and Shanks (2002) presented compelling evidence that participants generally use different strategies in making causal judgments, depending on the frequency with which judgments are collected across an experiment. These authors observed that causality judgments taken at different stages during training reflected the *momentary* predictive status of the CSs whereas ratings taken during a post-test reflected the *average* predictiveness taken across the *entire* training session. Lipp and Purkis (2006) suggest that a similar mechanism might operate in evaluative learning studies. Post-ratings might then reflect a stimulus' valence as averaged

[37] In traditional Pavlovian conditioning research (i.e., research on expectancy learning), it is well established that extinction is a context-dependent phenomenon (Bouton, 2004). In renewal experiments, for instance, a return of conditioned responding is observed when a context switch is made after the extinction phase (Bouton & Bolles, 1979). As explained in Part 2, it is still unclear whether evaluative learning is sensitive to contextual modulation. Lipp et al.'s (2003) critique on pre-test post-test EC studies rests on the assumption that this is the case.

across the entire experiment rather then its momentary valence and might therefore be misleading. Lipp and Purkis (2006) again refer to studies on the extinction-resistance of evaluative learning to illustrate their point. In their view, the post-experimental valence measures in these studies might reflect an integration of the valence the CS acquired during acquisition *and* extinction instead of the stimulus' current valence and therefore may wrongly suggest that evaluative learning is insensitive to extinction.

The foregoing arguments make clear that pre-test post-test designs have a number of important disadvantages and suggest that it might be preferable - or at least interesting - to (also) assess valence *during* rather than only *before* and *after* conditioning (i.e., online). In the previous section, an overview was given of the most commonly used valence measures in evaluation research. Several of these measures can be administered online, during evaluative learning. Self-report ratings, for instance, can and have been used to assess changes in evaluative learning in an online manner (e.g., Experiment 4 in Part 2; Lipp et al., 2003; Blechert et al., 2008). Online self-report ratings, however, suffer from the same disadvantage as pre/post ratings: they are rather susceptible to demand effects. The opportunity for demand effects might even be increased as online ratings are likely to draw participants' attention to the CS-US contingencies. The above described physiological and neurological measures also have the potential to be recorded at the same time that learning occurs. Hitherto, however, these measures have not been widely used in evaluation research, probably because of their high costs in terms of required expertise, equipment and labor. As was noted above, indirect behavioral RT tasks like the APT and IAT have gained widespread use in the EC literature due to their easy application. The currently existing indirect RT tasks, however, do not lend themselves (well) to be integrated in an ongoing conditioning procedure. In this second line of research, we therefore focused on the development of a variant of the APT that *can* be administered online, *during* evaluative learning. This 'online'[38] version of the APT circumvents the previously described problems that are associated with pre-test post-test measures, is easy in use and has the potential of providing more insight into the course of evaluative learning. Therefore, in our view, this measure fills an important gap in the literature and can help advance our understanding of (the properties of) evaluative learning. In the following section we will detail how this online APT was developed.

Assessing valence indirectly and online

In our search to develop an indirect and online behavioral valence measure, we were inspired by the work of Dawson, Beers, Schell, and Kelly (1982). These authors developed a secondary probe RT task to measure the operation of cognitive processes during human classical conditioning. Their use of a dual task technique is based on the assumption that humans' cognitive capacity is limited. The amount of cognitive processing invested in a certain task (referred to as the 'primary' task) therefore can be estimated by monitoring performance on a secondary task that is presented concurrently with this primary task. Deterioration of performance on the secondary task reflects the degree to which, and the time at which, the primary task requires limited-capacity processing resources. In the experiments of Dawson et al. (1982), participants' primary task was to pay attention to a differential classical conditioning procedure in which one CS (the CS+) was always followed by an electrocutaneous stimulus, while another CS (the CS-) was never followed by this US. Participants' secondary task was to respond as quickly as possible to tone probes that were presented at different moments during the CSs by pressing a response key. Dawson et al. (1982) found reaction times to be slower to tone probes presented during the CS+ than to probes presented during the CS-, which suggests greater capacity allocation to the reinforced stimulus. This effect was observed for probes presented at different timings (e.g., 300, 500, 3500 and 6500 ms after CS onset)

[38] In the present and following chapters we use the term 'online measure' to refer to a measure that has the *potential* to be administered during evaluative learning. It is clear that the qualification 'online' does not refer to a structural feature of a measure, but relates to how it is used (e.g., startle responses can be measured throughout conditioning, but can also be recorded in an offline manner, for instance, before and after a conditioning phase, e.g., Steidl & Yeomans, 2001).

within the seven seconds during CS intervals, but was most pronounced for early (i.e., 300 ms after CS onset) and late (i.e., 6500 ms) probes. According to the authors, the peak in capacity allocation shortly after CS+ onset reflects participants' processing of the US-signaling significance of this stimulus, while the late peak reflects their anticipation and preparation for the US. Since this first study, several other authors have applied the secondary probe RT task developed by Dawson et al. (1982) to track *expectancy learning* online in conditioning procedures (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Hermans, Dirikx, Vansteenwegen, Baeyens, Van den Bergh, & Eelen, 2005; Lipp, Siddle, & Dall, 1993).

We reasoned that an affective variant of this measure could be created by replacing the neutral tone probes with affect-laden visual probes (i.e., positive and negative icons) that have to be categorized by the participants. The in this way obtained task can actually be conceptualized as a variant of the APT with the CSs of the conditioning procedure serving as primes and the visual probes as targets. When a 'regular' APT is administered (as a pre- or post-test) in EC studies, the CSs also function as primes, but the prime stimulus is typically presented *shortly before* the target stimulus (see the studies in Part 2). In our adapted APT, however, the prime functions as a *background* stimulus against which the target stimuli appear, which allows our task to be integrated in an ongoing conditioning procedure. Like in a regular APT, we predict priming effects to occur in our task and thus expect participants to respond faster to probes that are evaluatively congruent with the CS than to probes that are evaluatively incongruent. Comparing participants' reaction times to positive versus negative target probes will then allow us to infer the evaluative value of the primes/CSs.

In the following chapters, we present three experiments in which we describe the set-up of our online variant of the APT in more detail and in which we tested its capacity to track (changes in) valence (during conditioning).

References

Anderson, A. K., & Sobel, N. (2003). Dissociating intensity from valence as sensory inputs to emotion. *Neuron, 39*(4), 581-583.

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*(2), 380-393.

Baeyens, F., Eelen, P., Crombez, G., & Van den Bergh, O. (1992). Human evaluative conditioning - Acquisition trials, presentation schedule, evaluative style and contingency awareness. *Behaviour Research and Therapy, 30*(2), 133-142.

Bar-Anan, Y., Nosek, B. A., & Vianello, M. (2009). The sorting paired features task a measure of association strengths. *Experimental Psychology, 56*(5), 329-343.

Bargh, J. A., Chaiken, S., Govender, R., & Pratto, F. (1992). The generality of the automatic attitude activation effect. *Journal of Personality and Social Psychology, 62*(6), 893-912.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*(3), 242-261.

Blanton, H., Jaccard, J., Gonzales, P. M., & Christie, C. (2006). Decoding the implicit association test: Implications for criterion prediction. *Journal of Experimental Social Psychology, 42*(2), 192-212.

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation, 39*(1), 58-70.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*(5), 485-494.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10*(4), 445-466.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion, 1,* 276-298.

Brunel, F. F., Tietje, B. C., & Greenwald, A. G. (2004). Is the implicit association test a valid and valuable measure of implicit consumer social cognition? *Journal of Consumer Psychology, 14*(4), 385-404.

Cacioppo, J. T., Crites, S. L., Berntson, G. G., & Coles, M. G. H. (1993). If attitudes affect how stimuli are processed, should they not affect the event-related brain potential. *Psychological Science, 4*(2), 108-112.

Cacioppo, J. T., Crites, S. L., & Gardner, W. L. (1996). Attitudes to the right: Evaluative processing is associated with lateralized late positive event-related brain potentials. *Personality and Social Psychology Bulletin, 22*(12), 1205-1219.

Cacioppo, J. T., Crites, S. L., Gardner, W. L., & Berntson, G. G. (1994). Bioelectrical echoes from evaluative categorizations: A late positive brain potential that varies as a function of trait negativity and extremity. *Journal of Personality and Social Psychology, 67*(1), 115-125.

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45*(1), 16-28.

Collins, A. M., & Loftus, E. F. (1975). Spreading activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Collins, D. J., & Shanks, D. R. (2002). Momentary and integrative response strategies in causal judgment. *Memory & Cognition, 30*(7), 1138-1147.

Cunningham, W. A., Raye, C. L., & Johnson, M. K. (2004). Implicit and explicit evaluation: fMRI correlates of valence, emotional intensity, and control in the processing of attitudes. *Journal of Cognitive Neuroscience, 16*(10), 1717-1729.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33*, 103-111.

Dawson, M. E., Beers, J. R., Schell, A. M., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical-conditioning. *Journal of Experimental Psychology-General, 111*(3), 273-295.

De Houwer, J. (2003a). The extrinsic affective simon task. *Experimental Psychology, 50*(2), 77-85.

De Houwer, J. (2003b). A structural analysis of indirect measures of attitudes. In J. Musch & K.C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219-244). Mahwah, NJ: Lawrence Erlbaum.

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11-28). Thousand Oaks, CA: Sage Publishers.

De Houwer, J. (2009). Comparing measures of attitudes at the procedural and functional level. In R. Petty, R. H. Fazio & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures*. Erlbaum.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*(2), 161-174.

De Houwer, J., & Eelen, P. (1998). An affective variant of the simon paradigm. *Cognition & Emotion, 12*(1), 45-61.

De Houwer, J., Hermans, D., & Eelen, P. (1998). Affective and identity priming with episodically associated stimuli. *Cognition and Emotion, 12*, 145-169.

De Houwer, J., & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies (pp. 179-194)*. New York: Guilford Press.

De Houwer, J., & Moors, A. (in press). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications*. New York, NY: Guilford Press.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin, 135*(3), 347-368.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853-869.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*(2), 245-268.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*(5), 549-554.

Dovidio, J. F., & Fazio, R. H. (1992). New technologies for the direct and indirect assessment of attitudes. In J. Tanur (Ed.), *Questions about survey questions: Meaning, memory, attitudes, and social interaction* (pp. 204-237). New York: Russell Sage Foundation.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and use. *Annual Review of Psychology, 54*, 297-327.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the Automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*(2), 229-238.

Fazio, R. H., & Williams, C. J. (1986). Attitude accessibility as a moderator of the attitude-perception and attitude-behavior relations - An investigation of the 1984 presidential-election. *Journal of Personality and Social Psychology, 51*(3), 505-514.

Ferguson, M. J., & Bargh, J. A. (2007). Beyond the attitude object: Automatic attitudes spring from object-centered-contexts. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Progress and controversies* (pp. 216-246). New York: Guilford Press.

Garavan, H., Pendergrass, J. C., Ross, T. J., Stein, E. A., & Risinger, R. C. (2001). Amygdala response to both positively and negatively valenced stimuli. *Neuroreport, 12*(12), 2779-2783.

Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgement and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology, 3*, 51-64.

Greenwald, A. G., & Farnham, S. D. (2000). Using the Implicit Association Test to measure self-esteem and self-concept. *Journal of Personality and Social Psychology, 79*(6), 1022-1038.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*(2), 197-216.

Grillon, C., & Baas, J. (2003). A review of the modulation of the startle reflex by affective states and its application in psychiatry. *Clinical Neurophysiology, 114*(9), 1557-1579.

Hamann, S., & Mao, H. (2002). Positive and negative emotional verbal stimuli elicit activity in the left amygdala. *Neuroreport, 13*(1), 15-19.

Hamann, S. B., Ely, T. D., Hoffman, J. M., & Kilts, C. D. (2002). Ecstasy and agony: activation of the human amygdala in positive and negative emotion. *Psychological Science, 13*(2), 135-141.

Hermann, C., Ziegler, S., Birbaumer, N., & Flor, H. (2000). Pavlovian aversive and appetitive odor conditioning in humans: Subjective, peripheral, and electrocortical changes. *Experimental Brain Research, 132,* 203-215.

Hermans, D., De Houwer, J., & Eelen, P. (1994). The Affective Priming Effect - Automatic activation of evaluative information in memory. *Cognition & Emotion, 8*(6), 515-533.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van Den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*(4), 533-551.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (in press). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*.

Isenberg, N., Silbersweig, D., Engelien, A., Emmerich, S., Malavade, K., Beattie, B., Leon, A. C., & Stern, E. (1999). Linguistic threat activates the human amygdala. *Proceedings of the National Academy of Sciences of the United States of America, 96*(18), 10456-10459.

Ito, T. A., & Cacioppo, J. T. (2000). Electrophysiological evidence of implicit and explicit categorization processes. *Journal of Experimental Social Psychology, 36*(6), 660-676.

Ito, T. A., & Cacioppo, J. T. (2007). Attitudes as mental and neural states of readiness: Using physiological measures to study implicit attitudes. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes* (pp. 125-158). New York: Guilford Press.

Karpinski, A., & Steinman, R. B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of Personality and Social Psychology, 91*(1), 16-32.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 7-49). Mahwah, NJ: Erlbaum.

Klauer, K. C., Musch, J., & Eder, A. B. (2005). Priming of semantic classifications: Late and response related, or earlier and more central? *Psychonomic Bulletin & Review, 12*(5), 897-903.

Klucken, T., Kagerer, S., Schweckendiek, J., Tabbert, K., Vaitl, D., & Stark, R. (2009). Neural, electrodermal and behavioral response patterns in contingency aware and unaware subjects during a picture-picture conditioning paradigm. *Neuroscience, 158*(2), 721-731.

Lang, P. J. (1995). The emotion probe - Studies of motivation and attention. *American Psychologist, 50*(5), 372-385.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*(3), 377-395.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1992). A motivational analysis of emotion - Reflex cortex connections. *Psychological Science, 3*(1), 44-49.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at Pictures - Affective, Facial, Visceral, and Behavioral Reactions. *Psychophysiology, 30*, 261-273.

Levenson, R. W., & Ekman, P. (2002). Difficulty does not account for emotion-specific heart rate changes in the directed facial action task. *Psychophysiology, 39*(3), 397-405.

Levey, A. B., & Martin, I. (1975). Classical-conditioning of human evaluative responses. *Behaviour Research and Therapy, 13*(4), 221-226.

Liberzon, I., Phan, K. L., Decker, L. R., & Taylor, S. F. (2003). Extended amygdala and emotional salience: A pet activation study of positive and negative affect. *Neuropsychopharmacology, 28*(4), 726-733.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 1-55.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative learning in human pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 431-440.

Lipp, O. V., Siddle, D. A. T., & Dall, P. J. (1993). Effects of miscuing on pavlovian conditioned responding and on probe reaction-time. *Australian Journal of Psychology, 45*(3), 161-167.

McDaniel, M. J., Beier, M. E., Perkins, A., Goggin, S., & Frankel, B. (2009). An assessment of the fakeability of explicit and implicit personality measures. *Journal of Research in Personality, 43,* 682-685.

Meiran, N., Chorev, Z., & Sapir, A. (2000). Component processes in task switching. *Cognitive Psychology, 41*(3), 211-253.

Mierke, J., & Klauer, K. C. (2001). Implicit association measurement with the IAT: Evidence for effects of executive control processes. *Zeitschrift Fur Experimentelle Psychologie, 48*(2), 107-122.

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85*(6), 1180-1192.

Mogg, K., Bradley, B. P., Field, M., & De Houwer, J. (2003). Eye movements to smoking-related pictures in smokers: Relationship between attentional biases and implicit and explicit measures of stimulus valence. *Addiction, 98,* 825-836.

Morris, J. S., Frith, C. D., Perrett, D. I., Rowland, D., Young, A. W., Calder, A. J., & Dolan, R. J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature, 383*(6603), 812-815.

Nosek, B. A., & Banaji, M. R. (2001). The go/no-go association task. *Social Cognition, 19*(6), 625-666.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*(5), 413-417.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning.* Chicago: University of Illinois Press.

Paradiso, S., Johnson, D. L., Andreasen, N. C., O'leary, D. S., Watkins, G. L., Ponto, L. L. B., & Hichwa, R. D. (1999). Cerebral blood flow changes associated with attribution of emotional valence to pleasant, unpleasant, and neutral visual stimuli in a pet study of normal subjects. *American Journal of Psychiatry, 156*(10), 1618-1629.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*(3), 277-293.

Pratto, F., & John, O. P. (1991). Automatic vigilance - The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*(3), 380-391.

Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation, 32*, 84-99.

Rankin, R. E., & Campbell, D. T. (1955). Galvanic skin response to Negro and White experimenters. *Journal of Abnormal and Social Psychology, 51*, 30-33.

Rottenberg, J., Kasch, K. L., Gross, J. J., & Gotlib, I. H. (2002). Sadness and amusement reactivity differentially predict concurrent and prospective functioning in major depressive disorder. *Emotion, 2*, 135-146.

Schimmack, U., & Crites, S. L. J. R. (2005). The Structure of affect. In D. Albarracín, B. T. Johnson & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 397-435). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Schwartz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist, 54*, 93-105.

Schwartz, N. (2008). Attitude measurements. In W. D. Crano & R. Prislin (Eds.), *Attitudes and attitude change* (pp. 41-60). New York: Psychology Press.

Schwartz, G. E., Brown, S. L., & Ahern, G. L. (1980). Facial muscle patterning and subjective experience during affective imagery - Sex-differences. *Psychophysiology, 17*(1), 75-82.

Small, D. M., Gregory, M. D., Mak, Y. E., Gitelman, D., Mesulam, M. M., & Parrish, T. (2003). Dissociation of neural representation of intensity and affective valuation in human gustation. *Neuron, 39*(4), 701-711.

Sriram, N., & Greenwald, A. G. (2009). The brief implicit association test. *Experimental Psychology, 56*(4), 283-294.

Steidl, S., Li, L., & Yeomans, J. S. (2001). Conditioned brain-stimulation reward attenuates the acoustic startle reflex in rats. *Behavioral Neuroscience, 115*(3), 710-717.

Tassinary, L. G., & Cacioppo, J. T. (1992). Unobservable facial actions and emotion. *Psychological Science, 3*(1), 28-33.

Tassinary, L. G., Cacioppo, J. T., & Geen, T. R. (1989). A psychometric study of surface electrode placements for facial electromyographic recording: The brow and cheek muscle regions. *Psychophysiology, 26*(1), 1-16.

Teachman, B. A. (2006). Pathological disgust: In the thoughts, not the eye, of the beholder. *Anxiety Stress and Coping, 19*(4), 335-351.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(1), 71-79.

Vrana, S. R., Spence, E. L., & Lang, P. J. (1988). The startle probe response - A new measure of emotion. *Journal of Abnormal Psychology, 97*(4), 487-491.

Whalen, P. J. (1998). Fear, vigilance, and ambiguity: Initial neuroimaging studies of the human amygdala. *Current Directions in Psychological Science, 7*(6), 177-188.

Wilbarger, J. L., McIntosh, D. N., & Winkielman, P. (2009). Startle modulation in autism: Positive affective stimuli enhance startle response. *Neuropsychologia, 47*(5), 1323-1331.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review, 107*(1), 101-126.

Wittenbrink, B. (2007). Measuring attitudes through priming. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes*. New York: Guilford Press.

experiment 1

Assessing valence indirectly and online: A pilot study

In the previous chapter we explained how - inspired by the work of Dawson, Beers, Schell, and Kelly (1982) - we developed a variant of the Affective Priming Task (APT) that can be administered online, during evaluative learning. In this chapter, a pilot study is presented in which we tested whether this task was capable of assessing the valence of normatively selected positive and negative IAPS pictures. Results indicated that our task succeeded in doing so, thereby supporting its validity.

Introduction

In the previous chapter, we discussed the potential benefits of a behavioral valence measure that can be administered online (i.e., during evaluative learning). We further described that in our search to develop such a measure, we were inspired by the work of Dawson et al. (1982). These authors measured the operation of cognitive processes in a differential fear conditioning procedure by asking participants to respond to tone probes that were presented during the CSs. We reasoned that an affective variant of this measure could be created by replacing the neutral tone probes with affect-laden (i.e., positive and negative) probes. We already mentioned that this affective version of Dawson et al.'s (1982) reaction time (RT) task can also be conceptualized as a variant of the APT with the CSs of the conditioning procedure serving as primes and the visual probes as targets. The main difference with the classic APT is that in our task the prime is not presented shortly before the target stimulus, but functions as a background stimulus against which the target stimuli appear.

In the present experiment, we tested the validity of our developed task by examining whether it was capable of assessing the valence of normatively selected positive and negative IAPS pictures (International Affective Picture System; Lang, Bradley, & Cuthbert, 2005). Participants were asked to categorize positive and negative probes (targets) that appeared at different timings *during* the presentation of these pictures (primes). Like in a regular APT, we expected participants to respond faster to congruent prime-probe pairs than to incongruent prime-probe pairs.

The current experiment was seen as a pilot for the study that is presented in the next chapter (Experiment 2), in which we wanted to test whether our task was able to assess valence *online*, *during* an evaluative conditioning procedure (i.e., the purpose it was developed for). The procedural features of the present experiment (e.g., prime duration: a CS duration of seven seconds is common in fear conditioning studies, see for instance Dawson et al., 1982; Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004) were therefore chosen in function of this planned study.

Method

Participants

The 26 participants (13 men, 13 women) were all volunteers. The majority of them were PhD students of the Department of Psychology. One participant was excluded from the study because of health problems which interfered with the task demands.

Materials

Four positive (*M* = 7.90, *SD* = 1.37) and four negative (*M* = 3.29, *SD* = 1.69) images, matched for arousal (M_{pos} = 4.55, SD_{pos} = 2.43, M_{neg} = 4.78, SD_{neg} = 2.20), were selected from the IAPS.[39] The pictures had a width of 512 pixels and a height of 384 pixels and were always presented against the black background of the computer monitor.

For the practice phase two neutral pictures - a picture of a key (*M* = 1.03, *SD* = 1.75) and a spoon (*M* = 0.47, *SD* = 0.90) - were selected on the basis of a rating study in which participants (*N* = 30) rated the affective connotation of 58 real life color pictures on a 11-point rating scale, ranging from -10 *(very negative)* to +10 *(very positive)*. The size of the practice pictures was identical to those selected from the IAPS.

The visual probe consisted of five positive (mouth turned upwards) or five negative (mouth turned downwards) black-and-white smiley faces that were presented for 1500 ms together with the IAPS pictures. One smiley face was presented in each corner and one in the middle of the picture (see Figure 11). Each smiley face had a width and height of 50 pixels.

The presentation of all stimuli, as well as the registration of all responses were controlled by an object-oriented, pool-based, real-time and millisecond accurate program (Affect 4.0), which was developed with C++ for the Windows platform (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010). Responses to the probes were recorded with a response box with two response buttons. An IBM compatible Pentium IV computer with 17 inch SVGA monitor was used to run the experiment.

---

[39] IAPS numbers: 1440 (Seal), 1710 (Puppies), 2311 (Mother), 5780 (Nature), 1111 (Snakes), 1270 (Roach), 9440 (Skulls), 2900.1 (Crying Boy).

Figure 11. Example of a positive (left) and negative (right) probe presented together with a positively (left) or negatively (right) valenced prime stimulus (original stimuli were color pictures from the IAPS, this figure shows pictures obtained from the internet with a similar content to the original stimuli).

Procedure

Participants were seated in front of the computer screen and were informed that different pictures would be presented. They were told that positive or negative smiley faces could appear during the presentation of these pictures and were instructed to categorize these as fast as possible by pressing the left button of the response box for negative smiley faces and the right button for positive smiley faces.

The experiment started with a practice phase of 12 trials during which two neutral pictures (key/spoon) were each presented six times for the duration of seven seconds. The order of the trials was randomized with the restriction that the first trial contained no visual probe. The intertrial interval was 1050 ms. Visual probes were presented for 1500 ms during four presentations of each picture. Two trials of each picture were presented without probe. To further increase unpredictability, probes could be presented between 300 and 1725 ms after picture onset ('early' probe), between 4075 and 5500 ms after onset ('late' probe) or at both moments.[40] For each picture in the practice phase there was one trial with an 'early' probe, one trial with a 'late' probe and two trials with an

[40] Probe timing was chosen so as to provide the participants with sufficient time to inspect the CSs and respond to the probes [e.g., the time gap between early (300-1750 ms) and late (4075-5500) probes was inserted to ensure that participants had time to prepare for a possible second probe after they had already responded to an early probe].

'early' and a 'late' probe. For both pictures, three of the six presented probes were positive and three were negative.

Subsequently, participants completed two test blocks of 64 trials. In both blocks, each IAPS picture was presented eight times and lasted seven seconds. The order of the trials was randomized with the restriction that the same IAPS picture could not be presented on more than two consecutive trials and that the first trial of each block was probe free. The intertrial interval was 1050 ms. In each block, probes were presented during six presentations of each IAPS picture; two trials of each picture were presented without probe. Similar to the practice phase, probes could be presented at different timings. For each IAPS picture in each block there was one trial with a positive 'early' probe, one trial with a negative 'early' probe, one trial with a positive 'late' probe, one trial with a negative 'late' probe, two trials containing an 'early' and a 'late' probe (of which two probes were positive and two were negative), and two trials without probe.

At the end of the experiment, participants were asked to rate the valence of the IAPS pictures on a 21-point rating scale ranging from -100 *(very unpleasant)* to +100 *(very pleasant).*

Results

Valence ratings

A t-test revealed a highly significant difference in the evaluative ratings of the intended positive and negative IAPS pictures, $t(24) = 14.22$, $p < .0001$. As expected, the intended positive IAPS pictures were rated far more positively ($M = 58.21$, $SD = 20.07$) than the intended negative IAPS pictures ($M = -52.80$, $SD = 26.59$).

Reaction time data

Data from trials on which no or an incorrect response was given were excluded from the analysis (5.44%). In addition, all response latencies shorter than 200 ms or longer than 1000 ms (see Dirikx et al., 2004) were excluded to reduce the influence of outlier responses (1.69%). The remaining data were analyzed with a 2 x 2 ANOVA with affective congruence (congruent: positive IAPS picture with positive probe, negative IAPS picture with negative probe / incongruent: positive

IAPS picture with negative probe, negative IAPS picture with positive probe) and block (1/2) as within-subjects variables. Since preliminary analyses showed no (main or interaction) effects of probe timing (early/late) and probe number (one/two), the RT data were pooled across these variables.

We expected participants to respond faster to congruent probes (i.e., probes that have the same valence as the prime stimulus) than to incongruent probes (i.e., probes that have an opposite valence). In line with expectation, the analysis of the RT data revealed a main effect of congruence, $F(1, 24) = 5.17$, $p < .05$, with faster responses to congruent probes ($M = 569$ ms, $SD = 65$) than to incongruent probes ($M = 580$, $SD = 65$). The main effect of block and the block by congruence interaction did not reach significance, with respectively $F(1, 24) = 1.08$, *ns* and $F < 1$.

Correlation between the reaction time data and evaluative ratings

For each IAPS picture, a 'positivity index' was calculated from the RT data by subtracting the mean response latency to positive probes from the mean response latency to negative probes presented during this picture. Higher scores on this variable are assumed to indicate a more positive picture evaluation. Supporting the convergent validity of our APT variant, a regression analysis controlling for the covariate 'subject', revealed a significant positive correlation between participants' ratings and RT indices, with $r = .15$, $p < .05$. Subject was added as a covariate to control the fact that each participant contributed multiple observations (i.e., for each participant eight rating scores and eight RT scores (one for each IAPS picture) were available). The positive correlation indicates that for each participant an increase in positivity in his/her ratings is associated with an increase in positivity in his/her RT scores.

Discussion

In the present experiment, we sought to establish the validity of our affective variant of Dawson et al.'s (1982) task by examining its sensitivity to the evaluation of items that are normatively regarded as favorable or unfavorable. Based on the RT data, we can conclude that our indirect measure was successful in grasping the valence of the IAPS pictures. Results showed that the affective valence of the IAPS pictures influenced participants' responses to the visual probes. As expected, participants responded faster when prime and probe were affectively congruent as compared to when they were incongruent. The convergent validity of our RT task was further supported by a significant positive correlation between participants' evaluative ratings and RT scores.

Structurally, our task is very similar to an APT. The main difference is that in our procedure the prime is not presented shortly before the target stimulus, but functions as a background stimulus against which the target stimuli appear. Like Dawson et al. (1982), we presented probes at different moments in the CS interval to keep them unpredictable for the participants and to allow multiple probe presentations - and hence the collection of more data - per trial. An important implication of this procedural difference is that the Stimulus Onset Asynchrony (SOA, i.e., the time interval between the beginning of the prime and the target) was variable in our task and rather long for probes presented at a late(r) time in the CS interval.[41] The fact that priming effects were observed at such long SOAs is remarkable because several studies found priming effects to disappear when the SOA exceeded 300 ms (e.g., Hermans, De Houwer, & Eelen, 2001; Klauer, Rossnagel, & Musch, 1997). This observation is typically seen as evidence for the fact that affective priming effects are grounded on fast-acting automatic processes (e.g., Hermans, Spruyt, & Eelen, 2003). Important to note, however, is that studies that have examined the influence of SOA length (e.g., Hermans et al., 2001; Klauer et al., 1997) have typically manipulated the onset asynchrony between prime and target by increasing or decreasing the interstimulus interval (ISI, the time interval between the offset of the prime and the onset of the target) while keeping prime duration constant and very short (i.e., 200 ms). Hence, in these studies SOA length (i.e., prime duration + ISI) and ISI length were always confounded. This is important, as a study by Hermans, Baeyens, and Eelen (1998) suggests that the length of the ISI rather than that of the SOA might be crucial for the occurrence of priming effects. In their study, Hermans et al. (1998) observed priming effects in an APT with a

[41] In the present experiment, SOAs varied between 300 and 5500 ms. Probe timing and SOA length covary in our procedure as the target stimuli were presented *during* the prime presentations.

long SOA (i.e., 10 sec) but short ISI (i.e., 0 ms).[42] According to these authors, their results indicate that a long SOA that is due to an increase in prime duration is not necessarily detrimental for priming effects. Hermans et al. (1998) suggest that a loss of priming effects will only occur for long SOAs that result from an increase in the ISI. In their opinion, the loss of effects in the latter case is caused by the fact that the activating influence of a prime gradually wears off after its offset. Similar to in the study of Hermans et al. (1998), the long SOAs in our experiment resulted from a long prime duration (i.e., 7 sec). The ISI on the other hand was always very short (i.e., 0 ms as targets appeared against the background of the prime stimulus). The fact that priming effects were obtained in our study therefore supports Hermans et al.'s (1998) hypothesis that ISI length rather than SOA length might be critical for priming effects. Additionally, Hermans et al.'s and our findings suggest that when the duration of the prime is increased, the valence of this stimulus remains activated over this time period. This finding is interesting as many authors would assume such activation to dissipate quickly over time (e.g., Fazio, 2001).

In this pilot experiment, we tested the adequacy of our RT task to assess stable pre-existing attitudes. In the following study, we wanted to incorporate our measure within an evaluative conditioning procedure to test whether it can be used to assess participants' attitudes *online*, *during* evaluative learning (the purpose this task was developed for). Therefore, we conducted an experiment in which we applied our task within a differential fear conditioning procedure. In this procedure, one CS (CS+) was consistently paired with an aversive US (i.e., an electrocutaneous stimulus) while another CS (CS-) was never paired with this US. Previous research has repeatedly shown that in such a design the CS+ typically acquires a negative connotation due to its association with the aversive US, while the CS- becomes a positive safety signal (e.g., Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006). Because of the controversy surrounding the extinction resistance of evaluative learning (see the comments of Lipp, Oughton, & Lelievre, 2003 and Lipp & Purkis, 2006 that were discussed

[42] The long SOA in this study thus resulted from a long (i.e., 10 seconds) prime presentation as SOA = prime duration + ISI.

in the previous chapter), we also added an extinction phase to the conditioning procedure in which both CSs were presented unreinforced.

References

Dawson, M. E., Beers, J. R., Schell, A. M., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical-conditioning. *Journal of Experimental Psychology - General, 111*, 273-295.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11,* 549-554.

Fazio, R. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15*, 115-141.

Hermans, D., Baeyens, F., & Eelen, P. (1998). Odors as affective processing context for word evaluation: A case of cross-modal affective priming. *Cognition & Emotion, 12,* 601-613.

Hermans, D., De Houwer, J., & Eelen, P. (2001). A time course analysis of the affective priming effect. *Cognition & Emotion, 15,* 143-165.

Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion, 17,* 83-99.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy, 40,* 217-234.

Klauer, K. C., Rossnagel, C., & Musch, J. (1997). List-context effects in evaluative priming. *Journal of Experimental Psychology - Learning Memory and Cognition, 23*(1), 246-255.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Digitized photographs, instruction manual and affective ratings. Technical Report A-6.* University of Florida, Gainesville, FL.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*, 219-239.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 431-440.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 71-79.

experiment 2

Assessing valence indirectly and online

In the present experiment, participants performed a reaction time task that was modeled after the affective priming procedure and was designed to track (changes in) the evaluative meaning of the conditioned stimuli (CSs) in a differential fear acquisition and subsequent extinction procedure indirectly and online. We asked participants to classify a visual probe that could appear during the reinforced CS and the unreinforced CS. The visual probe consisted of five positive or five negative smiley faces. Results indicated that the online task succeeded in tracking (shifts in) valence; superior performance was observed when CS and probe were congruent in affective meaning as compared to when they were incongruent. This congruency effect was not only observed at the end of the acquisition phase but also at the beginning and end of the extinction phase, which suggests that the acquired valence of the CSs survived extinction.

Published as: Kerkhof, I., Goesaert, E., Dirikx, T., Vansteenwegen, D., Baeyens, F., D'Hooge, R., & Hermans, D. (2009). Assessing valence indirectly and online. *Cognition & Emotion,* 23, 1615-1629.

Introduction

In the domain of attitude research a distinction is made between direct and indirect methods of attitude assessment. Direct attitude assessment relies on self-report methodology like questionnaires or interviews, whereas in indirect attitude assessment attitudes are inferred from behavior without directly asking respondents for their attitude. Indirect attitude measures are assumed to be less influenced by social desirability and intentional deception as compared to direct verbal reports and might even reflect attitudes that the respondent is unaware of (e.g., Asendorpf, Banse, & Mücke, 2002; but see Gawronski, LeBel, & Peters, 2007). The most frequently used indirect measures are the Affective Priming Task (APT) developed by Fazio and colleagues (Fazio, Sanbonmatsu, Powell, & Kardes, 1986) and Greenwald, McGhee, and Schwartz's (1998) Implicit Association Test (IAT). Often developed in the context of social psychology research, indirect measures are now commonly used in different areas of psychology.

Traditionally, indirect measures have been applied to assess attitudes considered to be highly stable and long lasting (Gawronski et al., 2007). Only more recently are these measures increasingly employed in studies that focus on newly acquired attitudes and the process of attitude formation and change. In a series of studies in the domain of aversive conditioning, for example, it was demonstrated that originally neutral stimuli that were predictive of an electrocutaneous stimulus not only acquired 'signal value' but also acquired a negative affective valence. In several studies this acquired negative affective valence was not only evidenced by data from evaluative rating scales, but also by data from an affective priming procedure (e.g., Hermans, Vansteenwegen, Crombez, Baeyens, & Eelen, 2002; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006). Also, in different studies on evaluative conditioning, indirect measures have been successfully used to assess newly conditioned attitudes (e.g., using the APT: Verhulst, Hermans, Baeyens, Spruyt, & Eelen, 2006; using the IAT: Olson & Fazio, 2001).

In these studies, indirect measures are typically administered in a session that precedes and/or follows the 'attitude formation phase' (pre-post measurements). This approach can have two important disadvantages. First, repeated administrations of an indirect measure are not problem-free and can lead to smaller effects (e.g., Greenwald, Nosek, & Banaji, 2003). Second, Lipp, Oughton and Lelievre (2003) pointed out that pre-post measurements could result in artifacts as learning effects can be sensitive to context changes. Several studies have documented shifts in indirectly measured attitudes based on variations in context, time or physiological state (e.g., Barden, Maddux, Petty, & Brewer, 2004). Indirect measures are often administered under conditions that differ from the attitude formation phase and even the mere transition of time between acquisition phase and measurement could be sufficient to constitute a context change (Bouton, 2004). Hence, it would be preferable to assess attitudes *during* the formation process (i.e., *online*). An online valence measurement can moreover provide more insight into how attitudes are formed and changed.

Various psychophysiological techniques have been put forward as online and indirect measure of attitudes, like facial EMG activity (e.g., Lang, Greenwald, Bradley, & Hamm, 1993) and the modulation of the acoustic startle reflex (e.g., Lang, Bradley, & Cuthbert, 1990). The main disadvantage of psychophysiological measures is, however, their cost in terms of required expertise, equipment and labor. Behavioral measurements like reaction time (RT) tasks are more easily applied using only the technology of personal computers. To our knowledge, no online indirect behavioral measurement of attitudes exists at present. Therefore in the current study, an adapted version of the APT was developed to indirectly track stimulus valence online.

In previous research on fear conditioning, a probe RT task (based on Dawson, Beers, Schell, & Kelly, 1982) was used as online measure of expectancy/fear (Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Hermans et al., 2005a). In these differential fear conditioning studies tone probes were presented during the presentation of a reinforced (CS+) and an unreinforced conditioned stimulus (CS-). Participants were asked to respond as fast as possible to the probes by pressing a key. Responses to tones presented during the CS+ were found to be slowed during fear acquisition, suggesting greater allocation of processing resources to this stimulus. In the present study an affective variant of this task was developed. Instead of tone probes affect-laden visual probes were presented during the CSs and participants were asked to evaluate the valence of these probes as fast as possible. We expected responses to be faster when the valence of the CS was congruent with the valence of the probe and slower when CS and probe valence were incongruent.

The task that was developed can be conceptualized as an online variant of the APT with the CSs serving as primes and the visual probes as targets. The main difference with the classic APT is that in our task the prime functions as a continuous background stimulus during which the targets are presented. A traditional finding in studies with the APT is that priming effects disappear when the Stimulus Onset Asynchrony (SOA, the time interval between the beginning of the prime and the target) is longer than 300 ms. This finding is typically seen as evidence for the fact that affective priming effects are grounded on fast-acting automatic processes (Hermans, Spruyt, & Eelen, 2003a). The way in which the SOA is typically manipulated is by increasing or decreasing the interstimulus interval (ISI); prime duration is generally kept very short (i.e., 200 ms). With increasingly longer ISIs (and hence longer SOAs) the possible activating effect of the prime gradually wears off. In the present experiment, however, the prime was continuously activated (as it functioned as a background stimulus) and is still present when the target is presented (ISI = 0). For this reason, we predicted to observe priming effects, even though a long SOA (due to the long prime duration) was applied. Similar to Hermans, Baeyens and Eelen (1998), who found affective priming effects when long lasting (i.e., 10 s) odors were used as primes, we believe the crucial variable for priming effects is not the SOA (prime duration + ISI) but the duration of the ISI.

In the present experiment, we applied the developed task to track shifts in the evaluative meaning of CSs in a differential fear acquisition and subsequent extinction procedure, with an electrocutaneous stimulus as unconditioned stimulus (US). In line with the results of previous studies, we expected the CS+ to gradually acquire a negative connotation throughout the acquisition phase (while the CS- receives a positive connotation). Based on previous studies from our lab that found no impact of an extinction procedure (i.e., repeated presentations of the CS alone after acquisition) on evaluative learning, little or no change in the acquired evaluative connotation of the CSs was expected during the extinction phase (Hermans, Crombez, Vansteenwegen, Baeyens, & Eelen, 2003b; Vansteenwegen et al., 2006).

Method

Participants

Fifty first-year psychology students (40 women) participated for partial fulfillment of course requirements. One participant was excluded from the study because of a technical error.

Materials

Two pictures of cookies served as CSs. One cookie was a yellow triangle; the other was a red moon shape. They were centered on a grey background and had a displayed diameter of 5-6 cm. The total picture had a width of 512 pixels and a height of 384 pixels. For the practice phase two neutral pictures (a picture of a key and of a spoon) were used.

The visual probe consisted of five positive (mouth turned upwards) or five negative (mouth turned downwards) black-and-white smiley faces that were presented for 1500 ms together with the CS-pictures. One smiley face was presented in each corner and one in the middle of the picture. Each had a width and height of 50 pixels.

A 2 ms electrocutaneous stimulus was used as US and was delivered by means of Fukuda standard Ag/AgCl electrodes (1.2 cm diameter). The electrodes were filled with electrode gel and were attached to the left wrist.

The presentation of all stimuli, as well as the registration of all responses were controlled by an object-oriented, pool-based, real-time and millisecond accurate program (Affect 4.0), which was developed with C++ for the Windows platform (Hermans, Clarysse, Baeyens, & Spruyt, 2005b). Responses to the probes were recorded with a response box with two response buttons. An IBM compatible Pentium IV computer with 17 inch SVGA monitor was used to run the experiment.

Procedure

The experiment consisted of a practice phase, an acquisition phase, an extinction phase and a post-experimental questionnaire. When participants entered the lab, the experiment was introduced as a study on human evaluations. Subsequently, they were informed about the use of electrocutaneous stimuli and of the possibility to decline to participate at any time during the experiment. The experiment started after the participant had given informed consent. Electrodes were attached to the left wrist and remained attached during the rest of the experiment. A work-up procedure was used to select the intensity of the US, which was set at a level that the participant described as 'unpleasant and demanding some effort to tolerate'.

Practice phase. The participant was seated at a distance of about 60 cm from the computer screen. The experiment started with a practice phase of 12 trials during which two neutral pictures were each presented six times for the duration of 7 s. The intertrial interval varied from 5500 to 6500 ms with a mean of 6000 ms. Visual probes were presented for the duration of 1500 ms during four presentations of each picture. To further increase the unpredictability of the probes, probes could be presented between 300 and 1725 ms after picture onset ('early' probe), between 4075 and 5500 ms after onset ('late' probe) or at both moments. For each picture in the practice phase there was one trial with only an 'early' probe, one with only a 'late' probe and two with an 'early' and a 'late' probe. For both pictures, three of the six presented probes were positive and three were negative. Participants were instructed to categorize the probes as fast as possible by pressing the left button of the response box for negative probes and the right button for positive probes.

Acquisition phase. After the practice phase, participants were informed that two new stimuli would be presented, one of which would sometimes be followed by the electrocutaneous stimulus, while the second stimulus would never be followed by this stimulus. Which of the two pictures of cookies that served as CS+ was counterbalanced across participants. In addition, participants were told that probes would again appear during the CS stimuli. They were instructed to remain attentive during the stimulus presentations and to categorize the probes as fast as possible.

The acquisition phase consisted of two blocks of 12 trials. These are referred to as the beginning/first half of acquisition and the end/second half of acquisition (see below). The CS+ and CS- were each presented six times in each block for 7 s. The order of these 12 trials was randomized, with the restriction that the first two trials always consisted of one CS+ and one CS- trial (order counterbalanced across participants) and that no stimulus could ever be presented on more than two successive trials. The intertrial interval varied from 5500 to 6500 ms with a mean of 6000 ms. The offset of the CS+ coincided with the onset of the US. The CS- was never followed by the US. Probes were presented for 1500 ms during four of the CS+ presentations and four of the CS- presentations. For each CS in each block there was one trial with only an 'early' probe, one with only a 'late' probe and two containing an 'early' and a 'late' probe. Both in the case of the CS+ and the CS-, three of the six presented probes were positive and three were negative.

Extinction phase. The extinction phase consisted of two series of 36 trials. These are referred to as the beginning/first half of extinction and the end/second half of extinction (see below). The presentation parameters for every 12 trials in the extinction phase were the same as for the acquisition blocks, with the only exception that no USs were presented. Participants were not informed in advance that the CS+ would no longer be followed by the US.

Questionnaire. After the electrodes were removed, participants were asked to identify the pictures used during the experiment by selecting the CS+ and CS- from a set of six pictures of cookies. Subsequently, they were asked to rate these two stimuli on a 21-point rating scale ranging from '-100/very unpleasant' to '+100/very pleasant'. Contingency awareness was assessed by asking participants to indicate for each CS whether it had been followed by the US during the first part of the experiment and how certain they were of this *(very uncertain, reasonably uncertain, reasonably certain, very certain)*. Additionally, participants rated the electrocutaneous stimulus for three characteristics. First, the (un)pleasantness of the US was rated on a 21-point scale with '-100/very unpleasant' and '+100/very pleasant' as extremes. The intensity of the US was assessed on a similar scale (ranging from 'light' over 'intense' to 'intolerable'). The extent to which the participants were startled by the US was appraised on a scale from ‘not at all’ over ‘moderately’ to ‘strongly’. The final questions addressed retrospective US-expectancy and fear. The experimenter explained that the study could be divided in four parts: *begin* (start of acquisition), *middle of first half* (end of acquisition), *middle of second half* (start of extinction) and *end* (end of extinction). Participants were asked for each of these moments to rate on a 10-point scale how much they had expected the US after each of both CSs. Similarly, they had to rate the level of fear they had experienced at these different moments when confronted with the CS-pictures.

Results

Data reduction and analyses

RT task trials on which no or an incorrect response was given, were discarded (2.55%). In addition, all response latencies shorter than 200 ms or longer than 1000 ms were excluded (3.08%). In order to obtain sufficient observations per cell and in order to compare the results of the RT task with the fear and expectancy ratings, RTs were averaged across halves of acquisition and halves of extinction. Because preliminary analyses showed no main or interaction effect of the probe timing (early/late) or probe number (one/two) variables, we averaged RTs across these variables in all subsequent analyses. The data were analyzed with a 2 x 2 x 2 ANOVA with phase (acquisition/extinction), moment (beginning/end) and affective congruence (congruent: CS+ with negative probe, CS- with positive probe / incongruent: CS+ with positive probe, CS- with negative probe) as within-subjects variables.

The data from the retrospective expectancy and fear ratings were analyzed using a 2 x 2 x 2 ANOVA with phase (acquisition/extinction), moment (beginning/end) and CS-type (CS+/CS-) as within-subjects variables.

RT data

The RT data indicated that the fear conditioning procedure led to significant shifts in the evaluation of the CSs (see Figure 12). A significant three-way interaction was observed between phase, moment and affective congruence, $F(1, 48) = 5.70$, $MSE = 919$, $p < .05$. In addition to this interaction, there were main effects of phase, $F(1, 48) = 7.48$, $MSE = 3127$, $p < .05$, and moment, F(1,

48) = 19.71, *MSE* = 1223, $p < .0001$, and a significant Phase x Moment interaction, $F(1, 48) = 19.71$, *MSE* = 1223, $p < .0001$. No other main effects or interactions were observed. The Phase x Moment x Affective congruence interaction was further examined using contrast analyses.

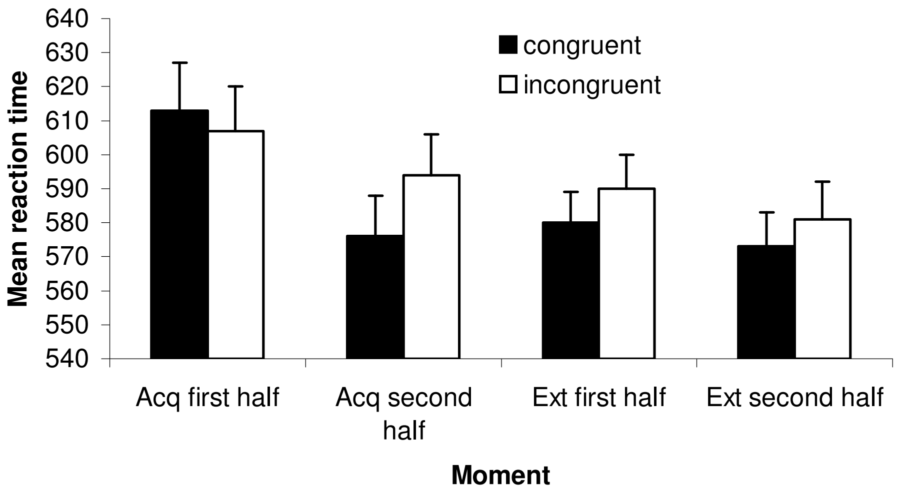


Figure 12. RTs in ms for congruent and incongruent CS-probe pairs for the first and second half of acquisition and extinction. Error bars represent standard errors.

<u>*Acquisition phase*</u>. Planned comparisons revealed no effect of affective congruence in the first half of acquisition, $F < 1$, M_{congr} = 613 ms, SD_{congr} = 96, $M_{incongr}$ = 604, $SD_{incongr}$ = 90. In the second part of acquisition, however, response latencies were shorter for congruent probes (M = 576, SD = 81) than for incongruent probes (M = 595, SD = 82), $F(1, 48) = 5.68$, *MSE* = 1517, $p < .05$. This suggests that during fear acquisition the CS+ acquired a negative valence and/or the CS- acquired a positive valence. The change in affective valence for the CS+ and CS- from the first to the second half of acquisition was evidenced by a significant Moment x Congruence interaction, $F(1, 48) = 4.47$, *MSE* = 2142, $p < .05$.

<u>*Extinction phase*</u>. For the first half of extinction, participants responded faster to congruent probes (M = 580, SD = 66) as compared to incongruent probes (M = 590, SD = 72), though this effect was only marginally significant, $F(1, 48) = 3.91$, *MSE* = 634, $p = .054$. For the second part of the extinction phase, a significant effect of congruence was observed, $F(1, 48) = 4.27$, *MSE* = 442, $p < .05$, with shorter response latencies for congruent probes (M = 573, *SD*

= 72) as compared to incongruent probes (*M* = 581, *SD* = 74), which suggests that the previously acquired valence of the CS+ and CS- (at least partly) survived extinction. The effect of affective congruence did not differ between both halves of the extinction phase. This was evidenced by the absence of a Moment x Congruence interaction, $F < 1$.

Acquisition phase versus extinction phase. As can be seen in Figure 12, the difference in ms between congruent and incongruent trials was larger during the second half of acquisition compared to the second half of extinction. Although this could be seen as evidence for some extinction of valence, the Phase x Congruence interaction was not statistically reliable, $F(1, 48) = 1.09$, *MSE* = 1118, *ns.* Hence, these findings also suggest that (at least part of) the valence acquired during acquisition outlasted extinction.

Post-experimental questionnaire data

Contingency awareness. Seven participants failed to correctly identify the CSs or to indicate which of the two was followed by the US. All participants were included in the analyses. Exclusion of the unaware participants did not affect the results.

Characteristics of the US. Participants rated the US as unpleasant (*M* = -59.18, *SD* = 19.13) and intense (*M* = 21.22, *SD* = 37.73), and indicated that they were moderately startled by the US (*M* = 25.51, *SD* = 43.78). Hence, the US was experienced as rather aversive.

Retrospective expectancy ratings. Mean US-expectancy ratings for both CSs for the four moments are presented in Figure 13 (upper panel). The crucial three-way interaction between phase, moment and CS-type was significant, $F(1, 48) = 35.43$, *MSE* = 4.08, $p < .0001$. In addition to this interaction, there were main effects of CS-type, $F(1, 48) = 122.30$, *MSE* = 9.14, $p < .0001$, phase, $F(1, 48) = 150.77$, *MSE* = 6.15, $p < .0001$, and moment, $F(1, 48) = 71.72$, *MSE* = 2.89, $p < .0001$ and a significant CS-type x Moment interaction, $F(1, 48) = 5.56$, *MSE* = 2.00, $p < .05$. None of the other interactions was significant. The three-way interaction between phase, moment, and CS-type was further examined using contrast analyses.

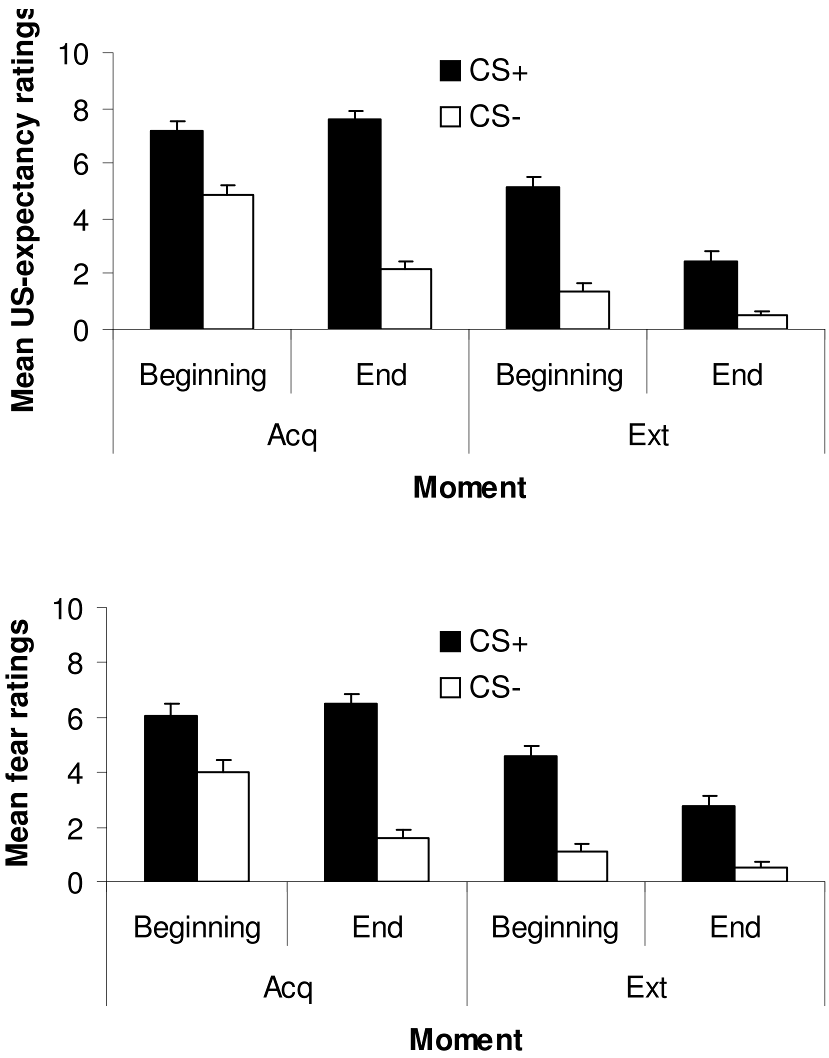


Figure 13. Mean US-expectancy (upper panel) and fear (lower panel) ratings for the CS+ and CS- at the beginning and end of acquisition and extinction. Error bars represent standard errors.

Contrasts revealed a significant differentiation in US-expectancy between the CS+ and CS- at the beginning of acquisition, $F(1, 48) = 23.17$, $MSE = 5.88$, $p < .0001$, as well as at the end of acquisition, $F(1, 48) = 114.36$, $MSE = 6.38$, $p < .0001$. The differentiation at the end of acquisition was significantly larger than at the beginning of acquisition, $F(1, 48) = 29.32$, $MSE = 4.02$, $p < .0001$, indicating successful acquisition of the contingencies. A significant differentiation was also observed at the beginning of extinction, $F(1, 48) = 51.91$, $MSE = 6.55$, $p < .0001$, which subsequently declined from the beginning to the end of extinction, $F(1, 48) = 18.36$, $MSE = 2.06$, $p < .0001$. At the end of extinction, there remained a small, but nevertheless significant difference in US-expectancy between the CS+ and CS-, $F(1, 48) = 28.14$, $MSE = 3.38$, $p < .0001$.

Retrospective fear ratings. A similar pattern was observed for the retrospective fear ratings (Figure 13, lower panel). Again, the crucial three-way interaction between phase, moment and CS-type was significant, $F(1, 48) = 37.76$, $MSE = 2.61$, $p < .0001$. In addition to this interaction, there were main effects of CS-type, $F(1, 48) = 103.01$, $MSE = 9.63$, $p < .0001$, phase, $F(1, 48) = 99.08$, $MSE = 5.16$, $p < .0001$, and moment, $F(1, 48) = 39.66$, $MSE = 3.04$, $p < .0001$, and a significant CS-type x Moment interaction, $F(1, 48) = 8.47$, $MSE = 1.72$, $p < .05$. No other interactions were obtained.

Contrasts revealed significantly higher fear ratings for the CS+ as compared to the CS- at the beginning of acquisition, $F(1, 48) = 22.34$, $MSE = 4.89$, $p < .0001$, as well as at the end of acquisition, $F(1, 48) = 98.30$, $MSE = 5.59$, $p < .0001$. The differentiation at the end of acquisition was significantly larger than at the beginning of acquisition, $F(1, 48) = 28.40$, $MSE = 3.32$, $p < .0001$. A significant differentiation was also obtained at the beginning of extinction, $F(1, 48) = 56.17$, $MSE = 5.28$, $p < .0001$. This differentiation declined significantly from the beginning to the end of extinction, $F(1, 48) = 18.62$, $MSE = 1.00$, $p < .0001$. At the end of extinction, there remained a small, but nevertheless significant difference in fear ratings between the CS+ and CS-, $F(1, 48) = 39.70$, $MSE = 3.11$, $p < .0001$.

Results of the retrospective expectancy and fear ratings provide clear evidence for successful differential acquisition of expectancy and fear learning. The results further indicate a considerable decline in differentiation between the CS+ and CS- from the beginning to the end of extinction.

Valence ratings. The analysis of the valence ratings revealed a highly significant difference in the evaluative ratings of the CSs at the end of the experiment, $t(48) = 7.55$, $p < .0001$, with the CS+ ($M = -26.94$, $SD = 47.97$) rated more negatively than the CS- ($M = 40.41$, $SD = 37.75$). Hence, the results of the valence ratings corroborate the findings of the indirect measure. Moreover, a significant correlation was observed between the difference in valence ratings between the CSs after extinction on the one hand (the rating for the CS+ was subtracted from the rating for the CS-), and the difference between congruent and incongruent RTs in the last part of the extinction phase on the other hand, $r(47) = .38$, $p < .01$. No similar relationships were observed with the residual difference in US-expectancy or fear after extinction, with respectively $r(47) = .11$, *ns*, $r(47) = .21$, *ns.* A multiple regression analysis confirmed that only the difference in valence ratings after extinction had a significant contribution in predicting the RT data at the end of the extinction phase, with $B = .17$, $\beta = .36$, $t(45) = 2.34$, $p < .05$. There was no significant contribution of the residual difference in US-expectancy, $B = -.45$, $\beta = -.04$, $t(45) = -.21$, *ns* or fear, $B = .78$, $\beta = .07$, $t(45) = .31$, *ns.*

Comparison of evaluative learning (RT task) and expectancy learning (ratings)

In order to be able to compare evaluative learning and expectancy learning directly, a MANOVA analysis was carried out. Therefore, all data were first Z-transformed. Subsequently, difference scores were calculated for the RT data and the ratings. For the RT data, the Z-transformed RT for congruent trials was subtracted from the Z-transformed RT for incongruent trials. For the ratings on the other hand, the Z-transformed rating for the CS+ was subtracted from the rating for the CS-. The data were analyzed using a 3 x 2 x 2 MANOVA with measure (expectancy ratings/fear ratings/RT data), phase (acquisition/extinction) and moment (beginning/end) as within-subjects variables. The crucial three-way interaction between measure, phase, and moment proved to be significant, $F(1, 48) = 11.31$, $MSE = .34$, $p < .0001$. In addition to this interaction, there were main effects of measure, $F(2, 96) = 66.19$, $MSE = .78$, $p < .0001$ and moment, $F(1, 48) = 9.28$, $MSE = .63$, $p < .01$, and a significant Phase x Moment interaction, $F(1, 48) = 46.92$, $MSE = .79$, $p < .0001$. No other main effects or interactions were observed. The Measure x Phase x Moment interaction was further examined using contrast analyses.

Acquisition phase. Contrasts revealed a stronger increase in differentiation during the acquisition phase for the fear and expectancy ratings as compared to the indirect measure, with $F(1, 48) = 5.75$, $MSE = .66$ and $p < .05$ when the expectancy ratings were compared with the RT data and $F(1, 48) = 5.48$, $MSE = .56$ and $p < .05$ when the fear ratings were compared with the RT data. Hence, stronger acquisition effects were found in the ratings as compared to the RT data. Important to note is that, despite the fact that acquisition might have been stronger in the ratings, a significant acquisition effect was also observed in the RT data (see earlier).

Extinction phase. Contrast analyses revealed that the decline in differentiation during the extinction phase was significantly larger for the fear and expectancy ratings as compared to the indirect measure, with $F(1, 48) = 11.25$, $MSE = .27$ and $p < .01$ when the expectancy ratings were compared with the RT data and $F(1, 48) = 8.87$, $MSE = .18$, and $p < .01$ when the fear ratings were compared with the RT data. As indicated earlier, whereas extinction effects were observed for the rating data, no such effects were apparent in the data of the RT task.

Discussion

In the present experiment, participants performed a RT task that was designed to track (changes in) the evaluative meaning of the CSs in a differential fear acquisition and subsequent extinction procedure indirectly and online.

Results indicated that our online RT task succeeded in tracking the expected changes in valence of the CSs. The RT data of the acquisition phase evidenced a gradual shift in the affective meaning of the CSs throughout the acquisition phase, with no effect of affective congruence in the first part of the acquisition phase but superior performance in the second part for congruent CS-probe pairs as compared to incongruent CS-probe pairs. These results corroborate earlier studies that found that originally neutral stimuli that were predictive of an electrocutaneous stimulus acquired a negative affective valence (e.g., Hermans et al., 2002). The RT data of the extinction phase suggest that (at least part of) the acquired valence of the CS+ and CS- survived extinction; in both halves of the extinction phase participants responded faster to congruent probes as compared to incongruent probes. These findings are in line with previous studies from our lab that found no impact of an extinction procedure on evaluative learning (e.g., Vansteenwegen et al., 2006). To our knowledge, this is the first time valence was assessed indirectly during the attitude formation process.

Successful differential acquisition was also observed in ratings of fear and expectancy. This learned differentiation between the CS+ and CS- declined considerably from the beginning to the end of extinction. As indicated above, a similar effect was not observed in the online valence measure. Hence, our findings corroborate previous studies that found a differential impact of extinction on evaluative learning versus expectancy learning (e.g., Hermans et al., 2003b; Vansteenwegen et al., 2006). In these previous studies, valence was always assessed in sessions that preceded and followed the acquisition and extinction training. According to Lipp et al. (2003) this relative resistance to extinction might be accounted for by renewal of evaluative learning due to a context switch from acquisition to post-test. Importantly, the present results cannot be attributed to such a renewal effect (or demand effects), as valence was measured online (and indirectly). Even though we do not intend to draw definitive conclusions about whether or not valence is *resistant* to extinction on the basis of the present data, it is clear that our findings support the idea that expectancy learning is at least ‘relatively’ *more susceptible* to extinction than evaluative learning. This hypothesis was supported by the conducted MANOVA analysis that indicated a stronger decline in differentiation during the extinction phase for the fear and expectancy ratings as compared to the indirect measure. A similar effect, however, was observed during acquisition; a stronger increase in differentiation was observed for the ratings as compared to the RT task. An alternative explanation for the different extinction results that were observed for the rating data and RT data in the MANOVA might then be that the RT task is simply a less sensitive measure than the ratings. Additionally, it can be argued that strong acquisition effects create more room for strong extinction effects whereas weak acquisition effects allow only for weak extinction effects. Important to keep in mind is, however, that despite the fact that acquisition might have been stronger in the ratings, a significant acquisition effect was also observed in the RT data. This makes clear that the RT task is sensitive enough to capture changes in the valence of the CSs. Moreover, despite the fact that this implies that there was room for extinction to occur in the RT measure, no such effect was observed. An interesting question for future research remains, however, whether our findings can be replicated when both evaluative learning and expectancy learning are assessed through an online indirect RT measure. Expectancy learning could hereby for instance be assessed with the RT task of Dawson et al. (1982).

The expectancy and fear ratings can be criticized for an additional reason, namely for their retrospective nature. Because the expectancy and fear ratings were taken retrospectively, they might possibly be confounded by the retrieval of the participant's knowledge about the experiment and experimental demands at the moment of test. Hence, also for this reason it would be interesting if future research would additionally include an online (in)direct measure of expectancy learning (i.e., the Dawson et al. (1982) task, skin conductance or online ratings with dials).

Importantly, the validity of our adapted RT task was supported by a multiple regression analysis that showed that the valence ratings were the only significant predictor of the congruence effect in the RT data. This suggests that stimulus valence is indeed the crucial element that drives the congruence effect in the online RT task.

As indicated in the introduction, the task we developed is structurally similar to the APT. A traditional finding in studies with the APT is that priming effects disappear when the SOA is longer than 300 ms. The SOA is hereby typically manipulated by increasing or decreasing the ISI. With increasingly longer ISIs (and hence longer SOAs) the possible activating effect of the prime gradually wears off. From this perspective, the priming effects that were observed in this study are due to the continuous activation of the prime that is still present when the target is presented (even though the onset of the prime was some seconds before the onset of the target). Hence, similar to Hermans et al. (1998), who found affective priming effects when long lasting odors were used as primes, our results suggest that the crucial variable for priming effects is not the SOA (prime duration + ISI) but the duration of the ISI (which was zero both in our study and the Hermans et al. study).

Further research could focus on ways to improve and learn more about the online measure that was presented in this study. A possible limitation of the present set-up is, for instance, that only a limited number of data points were collected per block. Given that there was no influence of probe timing or number on the results, a good approach for further applications of our RT measure would be to present more trials with double (or perhaps even more) probes. Another limitation of the present research is that demand awareness and compliance were not assessed. It would be interesting for future research to include such measurement to investigate whether the results of the RT task can be influenced by awareness of the experimental hypotheses. Last, but not least, future research could investigate the application of the online measure in other paradigms that are used to induce attitude formation and change.

References

Asendorpf, J. B., Banse, R., & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380-393.

Barden, J., Maddux, W. W., Petty, R. E., & Brewer, M. B. (2004). Contextual moderation of racial bias: The impact of social roles on controlled and automatically activated attitudes. *Journal of Personality and Social Psychology, 87*, 5-22.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*, 485-494.

Dawson, M. E., Beers, J. R., Schell, A. M., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical-conditioning. *Journal of Experimental Psychology - General, 111*, 273-295.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*, 549-554.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229-238.

Gawronski, B., LeBel, E. P., & Peters, K. R. (2007). What do implicit measures tell us? Scrutinizing the validity of three common assumptions. *Perspectives on Psychological Science, 2*, 181-193.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85,* 197-216.

Hermans, D., Baeyens, F., & Eelen, P. (1998). Odors as affective processing context for word evaluation: A case of cross-modal affective priming. *Cognition & Emotion, 12,* 601-613.

Hermans, D., Clarysse, J., Baeyens, F., & Spruyt, A. (2005b). Affect (Version 4.0) [Computer software; retrieved from http://www.psy.kuleuven.ac.be/leerpsy/affect4]. University of Leuven, Belgium.

Hermans, D., Crombez, G., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2003b). Expectancy-learning and evaluative learning in human classical conditioning: Differential effects of extinction. In P. L. Gower (Ed.), *Psychology of Fear* (pp. 133-156). New York: Nova Science Publishers.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005a). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*, 533-551.

Hermans, D., Spruyt, A., & Eelen, P. (2003a). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion, 17,* 83-99.

Hermans, D., Vansteenwegen, D., Crombez, G., Baeyens, F., & Eelen, P. (2002). Expectancy-learning and evaluative learning in human classical conditioning: Affective priming as an indirect and unobtrusive measure of conditioned stimulus valence. *Behaviour Research and Therapy, 40*, 217-234.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97*, 377-395.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures - Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*, 261-273.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative learning in human Pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*, 219-239.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413-417.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 71-79.

Verhulst, F., Hermans, D., Baeyens, F., Spruyt, A., & Eelen, P. (2006). Determinants and predictive validity of direct and indirect measures of recently acquired food attitudes. *Appetite, 46*, 137-143.

Assessing valence and expectancy indirectly and online

The results of Experiment 2 indicated that our affective variant of the secondary probe RT task of Dawson, Beers, Schell, and Kelly (1982) was capable of assessing the valence of the CSs indirectly and online. As mentioned earlier, Dawson et al. (1982) originally constructed their secondary task technique to assess the operation of cognitive processes during human classical conditioning. They presented (neutral) tone probes during the CSs of a differential fear conditioning procedure and found participants to respond slower to probes presented during the reinforced CS than to probes presented during the unreinforced CS, indicating larger resource allocation during the CS+. Since this first study, several authors have applied this secondary probe RT task as an indirect measure of expectancy/fear (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Hermans, et al., 2005; Lipp, Siddle, & Dall, 1993). Dirikx et al. (2004), for instance, used the secondary probe technique in an experiment on reinstatement of fear in humans. Slower RTs to probes presented during the CS+ as compared to probes presented during the CS- were observed during acquisition, but not during extinction. The presentation of two US alone trials after extinction reinstated the differential RT slowing, whereas no such effect was observed in a control group that did not receive US alone trials.

The affective version we have developed from Dawson et al.'s (1982) secondary probe technique is procedurally still very similar to the original task, with the main difference being that the neutral probes were replaced by affect-laden ones. Therefore, we assumed our adapted task to still be sensitive to expectancy learning. More specifically, in Experiment 2, we expected to observe slower responses to (positive and negative) probes presented in the CS+ interval than to probes presented in the CS- interval during acquisition and we expected this effect to disappear again during extinction (as expectancy learning is sensitive to extinction).

In our view, much benefit could be obtained from an indirect RT task that is able to assess *both* evaluative learning and expectancy learning during a conditioning procedure. In the General Introduction of this dissertation we already mentioned that some authors consider evaluative learning and expectancy learning to be two distinct forms of classical conditioning (e.g., Baeyens & De Houwer, 1995; Baeyens, Eelen, & Crombez, 1995). Theoretically, the distinction between these two types of learning rests on their different functional characteristics. Several evaluative conditioning (EC) studies found EC to demonstrate a number of properties that differ from expectancy learning. For example, unlike expectancy learning, evaluative learning has been found to not depend on awareness of the CS-US pairings (e.g., Baeyens, Eelen, & Van den Bergh, 1990; Walther & Nagengast, 2006), and to be insensitive to occasion setting or modulation (e.g., Baeyens, Crombez, De Houwer, & Eelen, 1996; Baeyens, Hendrickx, Crombez, & Hermans, 1998), CS-US contingency (e.g., Baeyens, Hermans, & Eelen, 1993), and extinction (e.g., Diaz, Ruiz, & Baeyens, 2005; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006). Other authors, however, contest the distinction between evaluative and expectancy learning as two different learning processes within Pavlovian conditioning. In their opinion, the observed differences between both types of learning might be due to procedural factors rather than to true process differences (e.g., Blechert, Michael, Williams, Purkis, & Wilhelm, 2008; Lipp & Purkis, 2005). Evaluative learning and expectancy learning indeed have been studied in very dissimilar research designs. Besides parametric differences such as the number of trials, number of stimuli, stimulus presentation duration, and interstimulus and intertrial intervals, research on evaluative learning and expectancy learning have typically differed in two ways. First, there is an important difference in the type of USs used in both lines of research. EC studies have most often used rather unobtrusive stimuli such as positively and negatively valenced pictures. In contrast, studies on expectancy learning have generally utilized biologically significant stimuli such as an electrocutaneous stimulus, which is an *unconditional* stimulus in a strict sense: it elicits a similar and innate/unlearned negative response in all subjects. Second, the response systems on which the dependent variables are based that have been used in both lines of research are often quite different. Most demonstrations of evaluative learning, for instance, are based on verbal ratings of CS valence while expectancy learning is often indexed by nonverbal autonomous responses such as the skin conductance response. In this sense, the nature of the utilized US or dependent variable rather than the expectancy/evaluative dimension might be critically involved in the observed differences between both types of learning.[43]

To exclude the possibility that procedural factors like US-type (or number of stimuli, stimulus duration, etc.) are responsible for the observed differences between expectancy and evaluative learning, it is advisable to assess both types of learning within a similar, or even better, a single design (e.g., Vansteenwegen et al., 2006; Experiment 4 in Part 2). To avoid that the use of different dependent variables (which might tap into different response systems or might differ in assessment properties[44]) generates artificial differences between both types of learning, it might be preferable to assess evaluative learning and expectancy learning with comparable measures (also see Blechert et al., 2008). Finding comparable measures of evaluative learning and expectancy learning is, however, not obvious. We reasoned that if our adapted version of Dawson et al.'s (1982) secondary probe technique would still represent a valid measure of expectancy learning, it would provide a promising tool for comparing evaluative learning and expectancy learning in a fair manner.[45] The task can easily be integrated in an ongoing conditioning procedure and would allow the assessment of both types of learning in the same response domain.

In the higher reported analyses of Experiment 2, we compared participants' responses to congruent and incongruent probes as an index of CS valence. To check whether our affective version of Dawson et al.'s (1982) task was still sensitive to expectancy learning, we compared participants' response latencies to probes presented during the CS+ with their latencies to probes

[43] An example of the latter can be found in the higher reported MANOVA analysis of Experiment 2 in which we compared changes in fear/expectancy ratings with changes in our developed RT task. This MANOVA analysis revealed a stronger decline in differentiation during the extinction phase for the fear and expectancy ratings as compared to the RT measure. In the discussion section of Experiment 2, we already indicated that these results might not reflect a 'true' process difference between expectancy learning and evaluative learning, but might relate to the use of different dependent variables to assess both types of learning (e.g., the RT task might be a less sensitive measure than verbal ratings or the difference in results might be due to the fact that the RT measure was administered online while the expectancy/fear ratings were collected retrospectively).

[44] E.g., potential to be administered as pre/post versus online measure.

[45] This is moreover one of the reasons for choosing the secondary probe technique of Dawson et al. (1982) as a basis for the construction of our online indirect behavioral valence measure and explains why we tried to keep the procedure of our adapted version as close as possible to the original task.

presented during the CS-. Based on the results of Dawson et al. (1982) and other experiments that applied this secondary probe technique (e.g., Dirikx et al., 2004), we expected to observe slower responses to CS+ probes than to CS- probes during the acquisition phase, but not during the extinction phase.

Results

To test the hypothesis that our affective variant of Dawson et al.'s (1982) task is still sensitive to expectancy learning, we analyzed the RT data of Experiment 2 with a 2 x 2 x 2 ANOVA with phase (acquisition/extinction), moment (beginning/end) and CS-type (CS+/CS-) as within-subjects variables. The crucial phase by moment by CS-type interaction was not significant, $F < 1$. A closer look at the data revealed that participants' response latencies to probes presented during the CS+ as compared to probes presented during the CS- did not differ at any moment in time (see Table 7, all *p*'s between .17 and .76).

Table 7
Mean response latencies and standard deviations in ms as a function of CS-type (CS+/CS-), phase (Acquisition, Extinction) and moment (Beginning/End)

		CS+		**CS-**	
		M	***SD***	***M***	***SD***
Acq	**Beginning**	611	95	608	95
	End	583	80	588	85
Ext	**Beginning**	583	70	589	67
	End	574	71	580	76

The analysis did reveal main effects of phase and moment with $F(1, 48) = 7.89$, $p < .01$, and $F(1, 48) = 19.73$, $p < .0001$. Participants responded faster to probes presented during the extinction phase than to probes presented during the acquisition phase and faster to probes presented in the second as compared to the first half of each phase. These main effects were further qualified by a significant Phase x Moment interaction, $F(1, 48) = 4.29$, $p < .05$. In each phase, participants responded faster during the second as compared to the first half, but this effect was more pronounced for the acquisition phase. These main and interaction effects are likely due to practice effects.

All participants were included in the above reported analysis. In previous studies with the original Dawson et al. (1982) task, contingency aware participants were typically excluded (e.g., Dirikx et al., 2004) as awareness is assumed to be a necessary prerequisite for expectancy learning to occur. Exclusion of the seven unaware participants did, however, not affect the results of our analysis.

Discussion

To test whether our adapted version of Dawson et al.'s (1982) secondary probe technique was also sensitive to expectancy learning, we compared participants' response latencies to probes presented during the CS+ with their latencies to probes presented during the CS-. Results indicated that participants responded equally fast to CS+ and CS- probes, both during acquisition and extinction. Hence, we failed to observe the standard expectancy/fear effect - slower responses to CS+ probes as compared to CS- probes during acquisition[46] - that was obtained in several studies that applied the original Dawson et al. (1982) task.

It is not clear why this effect did not emerge in our RT task. A closer look at the original Dawson et al. (1982) study led us to the hypothesis that a possible explanation for our replication failure might be found in how the RT task was presented to and perceived by the participants. In the study of Dawson et al., the *primacy* of paying attention to the CSs and USs during conditioning was strongly emphasized to the participants (e.g., '*Your primary task now is to pay attention to the colored lights and the shocks. At the end of the experiment we will ask you to answer some questions about the lights and the shocks, so it is important that you pay close attention to them.'* p. 280). In the instructions in our study, on the other hand, much more emphasis was put on the RT task. It was stressed to the participants that they had to focus on the probes and categorize them as fast as possible. They were also asked to remain attentive during the stimulus presentations, but paying attention to the CS-US contingencies was not explicitly described as an important or primary task. A possible reason for the

[46] Since we found no acquisition effect, the extinction data are also little meaningful.

absence of an expectancy effect in our data might then be that the categorization task was perhaps no longer a secondary task in our study and detracted participants' attention/resources from predicting and anticipating the US during the CS+ presentations. It is clear from our awareness data and expectancy ratings that participants learned the CS-US contingencies in our experiment. This does not exclude the possibility, however, that participants were *cognitively less engaged* in predicting and anticipating the US during the experiment.

In the next chapter, we present a third experiment in which we replicated the design of the present study, but made some crucial changes to the procedure that aimed at promoting a predictive mindset during conditioning in participants. We again explored whether our task was able to track both participants' evaluations and expectancies.

References

Baeyens, F., Crombez, G., De Houwer, J., & Eelen, P. (1996). No evidence for modulation of evaluative flavor-flavor associations in humans. *Learning and Motivation, 27*(2), 200-241.

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical-conditioning - Reply. *Behaviour Research and Therapy, 33*(7), 825-831.

Baeyens, F., Eelen, P., & Crombez, G. (1995). Pavlovian associations are forever - On classical-conditioning and extinction. *Journal of Psychophysiology, 9*(2), 127-141.

Baeyens, F., Eelen, P., & Van den Bergh, O. (1990). Contingency awareness in evaluative conditioning - A case for unaware affective-evaluative learning. *Cognition & Emotion, 4*(1), 3-18.

Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither extended sequential nor simultaneous feature positive training result in modulation of evaluative flavor-flavor conditioning in humans. *Appetite, 31*(2), 185-204.

Baeyens, F., Hermans, D., & Eelen, P. (1993). The role of CS-US contingency in human evaluative conditioning. *Behaviour Research and Therapy, 31*(8), 731-737.

Blechert, J., Michael, T., Williams, S. L., Purkis, H. M., & Wilhelm, F. H. (2008). When two paradigms meet: Does evaluative learning extinguish in differential fear conditioning? *Learning and Motivation, 39*(1), 58-70.

Dawson, M. E., Beers, J. R., Schell, A. M., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical-conditioning. *Journal of Experimental Psychology - General, 111*(3), 273-295.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*(2), 245-268.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*(5), 549-554.

Hermans, D., Dirikx, T., Vansteenwegen, D., Baeyens, F., Van den Bergh, O., & Eelen, P. (2005). Reinstatement of fear responses in human aversive conditioning. *Behaviour Research and Therapy, 43*(4), 533-551.

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple pavlovian conditioning. *Cognition & Emotion, 19*(2), 269-282.

Lipp, O. V., Siddle, D. A. T., & Dall, P. J. (1993). Effects of miscuing on pavlovian conditioned responding and on probe reaction-time. *Australian Journal of Psychology, 45*(3), 161-167.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(1), 71-79.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(4), 454-459.

experiment 3

Assessing valence and expectancy indirectly and online: A second attempt

In Experiment 2, we found our affective version of the secondary probe reaction time (RT) task of Dawson, Beers, Schell, and Kelly (1982) to be capable of assessing the valence of the CSs in an ongoing conditioning procedure. Contrary to our predictions, however, our task was no longer sensitive to expectancy learning. In the foregoing chapter, we hypothesized that the latter finding might have been caused by the fact that our affective categorization task detracted participants' attention/resources from predicting and anticipating the US during the conditioning procedure. In the present chapter, we present a study in which we replicated the design of Experiment 2, but made some crucial methodological changes that aimed at inducing a predictive mindset in participants. Nevertheless, the results indicated that our task again was able to capture participants' evaluations but not their expectancies. Potential reasons for the latter finding are discussed.

Introduction

In the previous experiment, we found our affective variant of Dawson et al.'s (1982) secondary probe RT technique capable of assessing CS valence in an ongoing conditioning procedure. Contrary to our predictions, however, our task was no longer sensitive to expectancy learning. In the foregoing chapter, we hypothesized that the latter finding might have been caused by the fact that our RT task captured too much of participants' resources and reduced their cognitive engagement in anticipating the US during the CS presentations. Put differently: our task was perhaps no longer secondary in the previous study.

In the present study, we replicated the design of Experiment 2, but made some crucial methodological changes that aimed at making the predictive component of the conditioning procedure more salient and relevant for the participants. The latter was then hoped to increase their cognitive involvement in predicting and anticipating the US during the experiment.

A first change concerned the instructions that were given to the participants at the beginning of the conditioning phase. In the instructions of Experiment 2, no special emphasis was put on the importance of paying attention to the CS-US relations. Participants were simply asked to remain attentive during conditioning and to focus on categorizing the probes. In the present study, we tried to increase participants' attention for the predictive value of the CSs by informing them that their *primary* task was to pay attention to the CS-US contingencies as they would receive questions about the relationship between the CSs and US at the end of the experiment. Categorizing the probes, on the other hand, was described as a *secondary* task. A second change was that expectancy ratings were integrated in the conditioning procedure of the present experiment. After each trial, participants were asked to indicate on a scale ranging from -100 (*not at all)* to +100 (*certainly*) to what extent they had expected the US to occur after the presented CS. The inclusion of these trial-by-trial US-expectancy ratings was also aimed at promoting a predictive mindset during conditioning. To make sure, however, that the conditioning procedure was not affected by the inclusion of the online expectancy ratings, these ratings were administered in the intertrial interval (and hence actually assessed participants' expectancies retrospectively). A third alteration concerned the shock work-up procedure that was used to determine the intensity of the electrocutaneous stimulus for each individual participant. In Experiment 2, we asked participants to select a shock intensity that was 'unpleasant and demanding some effort to tolerate'. In the present experiment, participants were asked to rate the shocks that were administered during the work-up procedure on a scale ranging from 0 ('*You do not sense anything at all*') to 10 ('*This is the maximum tolerable pain for you in this experiment*') and were strongly encouraged to select the *highest* level they could tolerate (procedure adapted from Fonteyne, Vervliet, Hermans, Baeyens, & Vansteenwegen, 2009). By encouraging participants to select a truly aversive US, we hoped to enhance their fear for the US and therefore also their desire to anticipate (and prepare themselves for) this stimulus. For the same purpose, a threat manipulation was included in the experiment. Participants were told at the start of the acquisition phase that the intensity of the US could increase during this phase. In reality, the intensity of the electrocutaneous stimulus remained unchanged throughout the experiment.

Based on some unpublished studies in our lab that applied the original Dawson et al. (1982) task (Joos, Vansteenwegen, & Hermans, 2008; Kerkhof, 2008), we also changed the timing of the probes in the present experiment. Both studies applied the secondary probe technique of Dawson et al. (1982) in a differential fear conditioning procedure and found the traditional expectancy effect (i.e., slower responding to CS+ probes than to CS- probes) to occur only for late but not for early probes. Because of these findings we decided to delay the timing of the early probes in the present experiment to 4500 ms after CS onset. In order to collect more data, we further presented probes during *each* CS trial and increased the number of trials on which both an early and a late probe were presented.

In the present experiment, we again explored whether our adapted version of Dawson et al.'s (1982) task was able to track both participants' evaluations and expectancies during conditioning. In contrast to Experiment 2, the current experiment contained only an acquisition phase as slower RTs for CS+ probes as compared to CS- probes (indexing differential US-*expectancies*) are only expected during this phase and not during extinction. Like in

Experiment 2, we further predicted to observe faster responses to congruent probes than to incongruent probes (indexing differential CS *evaluations*).

Method

Participants

Thirty-four first-year psychology students (28 women) participated for partial fulfillment of course requirements. Five participants who failed to comply with task instructions were excluded from the analyses.

Materials

Like in Experiment 2, two pictures of cookies served as CSs. One cookie was a yellow triangle; the other was a red moon shape. They were centered on a grey background and had a displayed diameter of 5-6 cm. The total picture had a width of 512 pixels and a height of 384 pixels. For the practice phase two neutral pictures (a picture of a key and of a spoon) were used.

The visual probe consisted of one positive (mouth turned upwards) or one negative (mouth turned downwards) black-and-white smiley face that was presented for 1000 ms together with the CS-pictures. This icon appeared in the middle of the CS-picture and had a width and height of 50 pixels.

A 2 ms electrocutaneous stimulus was used as US and was delivered by means of Fukuda standard Ag/AgCl electrodes (1.2 cm diameter). The electrodes were filled with electrode gel and were attached to the left wrist.

The presentation of all stimuli, as well as the registration of all responses were controlled by an object-oriented, pool-based, real-time and millisecond accurate program (Affect 4.0), which was developed with C++ for the Windows platform (Spruyt, Clarysse, Vansteenwegen, Baeyens, & Hermans, 2010). Responses to the probes were recorded with a response box with two response buttons. An IBM compatible Pentium IV computer with 17 inch SVGA monitor was used to run the experiment.

Procedure

When participants entered the lab, the experiment was introduced as a study on human evaluations. Subsequently, they were informed about the use of electrocutaneous stimuli and of the possibility to decline to participate at any time during the experiment. The experiment started after the participant had given informed consent. Electrodes were attached to the left wrist and remained attached during the rest of the experiment. A work-up procedure adopted from Fonteyne et al. (2009) was used to select the intensity of the US. The participant was told that during the work-up phase electrocutaneous stimuli would be administered with increasing intensity, starting with a stimulus that was very low in intensity. Participants were asked to rate the administered shocks on a scale from 0 to 10 where '0' meant: '*You do not sense anything at all*', '1' meant: '*You feel something but it is not painful, it is just a sensation*', '2' meant: '*It starts to be painful but it is still a very small pain*', up to '10' which meant: '*This is the maximum tolerable pain for you in this experiment'*. When reaching the highest level tolerated by the participant, he/she was asked to rate the unpleasantness of the chosen shock on a scale ranging from -100 (*very unpleasant)* to +100 (*very pleasant*). After this, the actual experiment started. It consisted of three consecutive phases: a practice phase, an acquisition phase and a post-experimental questionnaire.

Practice phase. The participant was seated at a distance of about 60 cm from the computer screen. The practice phase consisted of eight trials during which two neutral pictures were each presented four times for the duration of 7 s. The intertrial interval varied from 2000 to 3000 ms with a mean of 2500 ms. Visual probes were presented for the duration of 1000 ms and could appear 4500 ms after picture onset ('early' probe[47]), 6000 ms after picture onset ('late' probe) or at both moments. For each picture in the practice phase there was one trial with an 'early' probe, one with a 'late' probe and two with an 'early' and a 'late' probe. For both pictures, three of the six presented probes were positive and three were negative. Participants were instructed to categorize the probes as fast as possible by pressing the left button of the response box for negative probes and the right button for positive probes.

Acquisition phase. After the practice phase, participants were informed that two new stimuli would be presented, one of which would always be followed

[47] We use the term 'early' to refer to probes that appeared at the *earliest* timing used in the present experiment (even though these probes were actually not delivered early in the CS interval).

by the electrocutaneous stimulus, while the second stimulus would never be followed by this US. They were further told that the intensity of the shocks could increase during the experiment. This instruction was aimed at increasing participants' fear and therefore also their desire to anticipate the electrocutaneous stimuli. In reality, the intensity of the electrocutaneous stimulus remained unchanged throughout the experiment. Which of the two pictures of cookies that served as CS+ was counterbalanced across participants.

Participants were told that they had to perform two tasks. They were explained that their first and primary task was to pay attention to the CSs, USs, and their relationship during conditioning. They were also informed that they would receive questions about the CS-US relations at the end of the experiment. The importance of this task was further stressed by telling the participants that after each trial they would be asked to indicate on a scale ranging from -100 (*not at all*) to +100 (*certainly*) to what extent they had expected the US to occur. Subsequently, the participants were told that their second task consisted of categorizing the icons that would appear during the picture presentations.

The acquisition phase consisted of three blocks of 12 trials. In each block, the CS+ and CS- were each presented six times for a duration of 7 s. The order of these 12 trials was randomized, with the restriction that the first two trials always consisted of one CS+ and one CS- trial (order counterbalanced across participants) and that no stimulus could ever be presented on more than two successive trials. The intertrial interval varied from 5500 to 6500 ms with a mean of 6000 ms. The offset of the CS+ coincided with the onset of the US. The CS- was never followed by the US. Probes were presented for 1000 ms during each CS presentation. For each CS in each block there was one trial with an 'early' probe, one with a 'late' probe and four containing an 'early' and a 'late' probe. Both in the case of the CS+ and the CS-, five of the ten presented probes were positive and five were negative. At a fixed time (1500 ms) after CS offset, a rating scale appeared on the screen on which participants were requested to indicate to what extent they had expected the US to occur after the previously presented CS. The scale ranged from -100 (*not at all*) to +100 (*certainly*). In order to assess possible US habituation effects, participants were asked to rate the (un)pleasantness of the US after each block on a scale ranging from -100 (*very unpleasant)* to +100 (*very pleasant*).

Post-experimental questionnaire. After the acquisition phase, participants were asked to identify the pictures used during the experiment by selecting the CS+ and CS- from a set of six pictures of cookies. Subsequently, they were asked to rate these two stimuli on a 21-point rating scale ranging from -100 (*very unpleasant*) to +100 (*very pleasant*). Contingency awareness was assessed by asking participants to indicate for each CS whether it had been followed by the US during the experiment and how certain they were of this *(very uncertain, reasonably uncertain, reasonably certain, very certain).* Additionally, participants rated the electrocutaneous stimulus for three characteristics. First, the (un)pleasantness of the US was rated (for the fifth time) on a scale with -100 (*very unpleasant*) and +100 (*very pleasant*) as extremes. The intensity of the US was assessed on a similar scale (ranging from '*light*' to '*intolerable*'). The extent to which the participants were startled by the US was appraised on a scale ranging from '*not at all*' (-100) to '*strongly'* (+100). The final questions addressed retrospective US-expectancy and fear. Participants were asked to rate on a 10-point scale how much they had expected the US after each of both CSs at the beginning and end of the experiment. Similarly, they had to rate the level of fear they had experienced at both moments when confronted with the CS-pictures.

Results

RT data

RT task trials on which no or an incorrect response was given, were discarded (3.79%). In addition, all response latencies shorter than 200 ms or longer than 1000 ms were excluded (4.66%). Because preliminary analyses showed no meaningful effects of the probe timing (early/late) or probe number (one/two) variables, we averaged RTs across these variables in all subsequent analyses.

Evaluative learning. We expected the CS+ to acquire a negative connotation throughout conditioning because of its association with the aversive US and the CS- to become positive (as it functions as a safety signal for the participants). Regarding the RT data, this implies the prediction of faster responses to congruent probes than to incongruent probes.

To test this prediction, the RT data were analyzed using a 3 x 2 repeated measures ANOVA with block (1/2/3) and affective congruence (congruent: CS+ with negative probe, CS- with positive probe / incongruent: CS+ with positive probe, CS- with negative probe) as within-subjects variables. The analysis revealed main effects of affective congruence and block, with $F(1, 28) = 5.01$, $p < .05$ and $F(2, 56) = 4.58$, $p < .05$. In general, participants responded faster to congruent probes ($M = 535$, $SD = 46$) than to incongruent probes ($M = 547$, $SD = 52$) and were faster in later blocks of the task ($M_{B1} = 553$, $SD_{B1} = 52$; $M_{B2} = 537$, $SD_{B2} = 54$; $M_{B3} = 534$, $SD_{B3} = 49$). Participants' mean response latencies to congruent and incongruent probes as a function of block are depicted in Figure 14. It is clear from this figure that in each block, participants responded faster to congruent probes than to incongruent probes. These data indicate that, as predicted, the CS+ acquired a negative meaning and/or the CS- became positive. As can be seen in Figure 14, there was a trend towards an increase in the effect of congruence over blocks. The Block x Affective congruence interaction was not significant, however, $F < 1$.

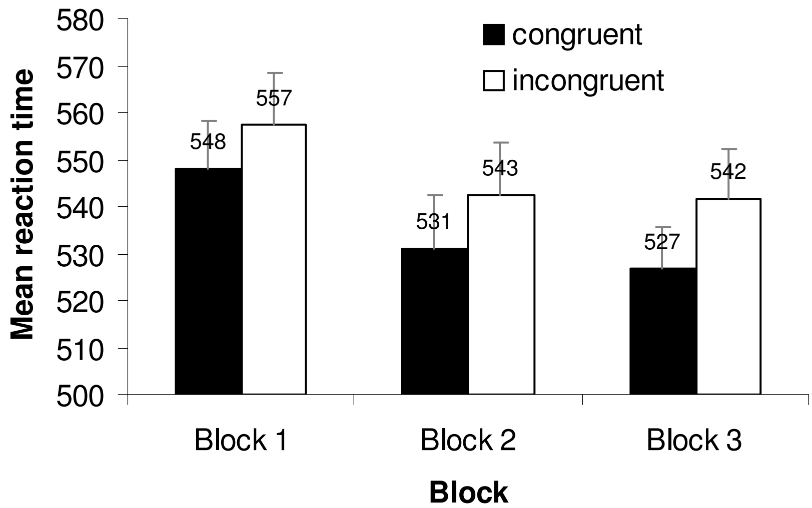


Figure 14. Reaction times in ms for congruent and incongruent CS-probe pairs as a function of block. Error bars represent standard errors.

<u>*Expectancy learning.*</u> Like in fear conditioning studies that applied the original Dawson et al. (1982) task, we expected participants to respond slower to probes presented during the CS+ than to probes presented during the CS-.

To test this hypothesis, the RT data were analyzed using a 3 x 2 repeated measures ANOVA with block (1/2/3) and CS-type (CS+/CS-) as within-subjects variables. This analysis only revealed a main effect of block, $F(2, 56) = 4.72$, $p < .05$, with participants responding faster over time (see earlier). The main effect of CS-type and the Block x CS-type interaction were not significant, with $F < 1$ and $F(2, 56) = 1.02$, *ns.* Participants responded equally fast to CS+ ($M = 541$, $SD = 53$) and CS- probes ($M = 542$, $SD = 47$). Hence, also in this experiment we failed to replicate the differential slowing effect that was found in several studies that applied the original Dawson et al. (1982) task.

Online expectancy ratings

We expected participants to gradually acquire the CS-US contingencies throughout the experiment. We predicted to observe little or no differentiation in the US-expectancy ratings for the CS+ and CS- at the beginning of acquisition, but significantly higher expectancy ratings for the CS+ as compared to the CS- towards the end of acquisition.

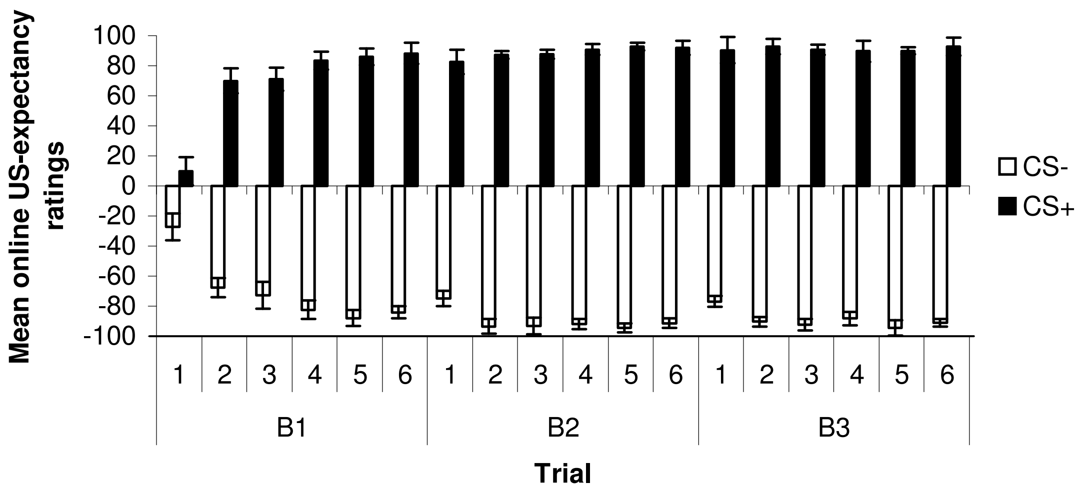


Figure 15. Mean online US-expectancy ratings as a function of trial and block. Error bars represent standard errors.

Participants' online expectancy ratings are depicted in Figure 15. It is clear from this figure that participants quickly differentiated the CS+ from the CS-, resulting in almost maximal US-expectancy associated with the CS+ and minimal expectancy with the CS- towards the end of the acquisition phase.

This pattern was confirmed in a 2 (CS-type: CS+/CS-) x 2 (Moment: first/last acquisition trial) ANOVA. Importantly, a significant Moment x CS-type interaction was obtained, $F(1, 28) = 151.81$, $p < .0001$, indicating a different pattern of results for the CS+ and CS-. As expected, there was a significant increase in US-expectancy for the CS+, $F(1, 28) = 74.35$, $p < .0001$ and a significant decrease for the CS-, $F(1, 28) = 38.57$, $p < .0001$. Significantly higher US-expectancy ratings for the CS+ as compared to the CS- were observed both at the beginning and at the end of the acquisition phase, $F(1, 28) = 16.02$, $p < .001$ and $F(1, 28) = 661.22$, $p < .0001$. The difference in expectancy ratings was, however, most pronounced at the end of the learning phase (Trial 1: $M_{CS+} = 9.66$, $SD_{CS+} = 9.12$; $M_{CS-} = -27.24$, $SD_{CS-} = 9.35$; Trial 18: $M_{CS+} = 92.76$, $SD_{CS+} = 2.62$; $M_{CS-} = -91.03$, $SD_{CS-} = 5.82$). Finally, the analysis revealed a main effect of CS-type, $F(1, 28) = 619.44$, $p < .0001$. In general (over trials), the CS+ elicited more US-expectancy than did the CS-.

US characteristics

Participants rated the (un)pleasantness of the US at five different moments: before the start of the experiment, after each acquisition block and at the end of the experiment. Participants' mean US ratings are displayed in Table 8. It is apparent from this table that participants experienced the electrocutaneous stimulus as very unpleasant at each test moment.

Table 8
Mean US ratings and standard deviations as a function of moment. Non-overlapping superscript letters indicate significant differences (p < .05) between rows

	M	***SD***
Before the experiment	-60.69^{a}	16.02
After acquisition block 1	$-64.14^{a,b}$	15.70
After acquisition block 2	-69.66^{c}	15.69
After acquisition block 3	-69.66^{c}	16.14
End of the experiment	$-67.93^{b,c}$	14.49

To check for possible shock habituation effects, the US ratings were analyzed using a one-way repeated measures ANOVA with moment (before the experiment, after acquisition block 1, after acquisition block 2, after acquisition block 3, at the end of the experiment) as a within-subjects variable. The analysis revealed a main effect of moment, $F(4, 112) = 6.14$, $p < .001$. The results of further contrast analyses are reported in Table 8. In general, there was a trend for more negative US ratings as the experiment progressed. Hence, we found no evidence for US habituation throughout the experiment.

Post-experimental questionnaire data

Contingency awareness. All participants were able to correctly identify the CSs and to indicate which CS was followed by the US.

Characteristics of the US. The (un)pleasantness ratings were reported above. Participants further rated the US as intense ($M = 44.14$, $SD = 17.01$) and indicated that they were moderately startled by this stimulus ($M = 42.76$, $SD = 24.77$). In sum, the US was experienced as rather aversive.

Valence ratings. As already mentioned, we expected the CS+ to become negative and the CS- to become positive throughout conditioning. A t-test performed on participants' post-acquisition valence ratings revealed a highly significant difference in the evaluation of the CSs, $t(29) = -9.69$, $p < .0001$. As predicted, the CS+ ($M = -54.14$, $SD = 40.05$) was rated more negatively than the CS- ($M = 58.62$, $SD = 33.78$). This finding corroborates the results of our (affective) RT data.

Retrospective expectancy ratings. Like for the online expectancy ratings, we expected to find a learning effect in the retrospective expectancy ratings. We expected to observe little or no differentiation in the US-expectancy ratings for the CS+ and CS- at the beginning of acquisition, but significantly higher expectancy ratings for the CS+ as compared to the CS- at the end of acquisition. Participants' mean retrospective US expectancy ratings are depicted in Figure 16 (left panel).

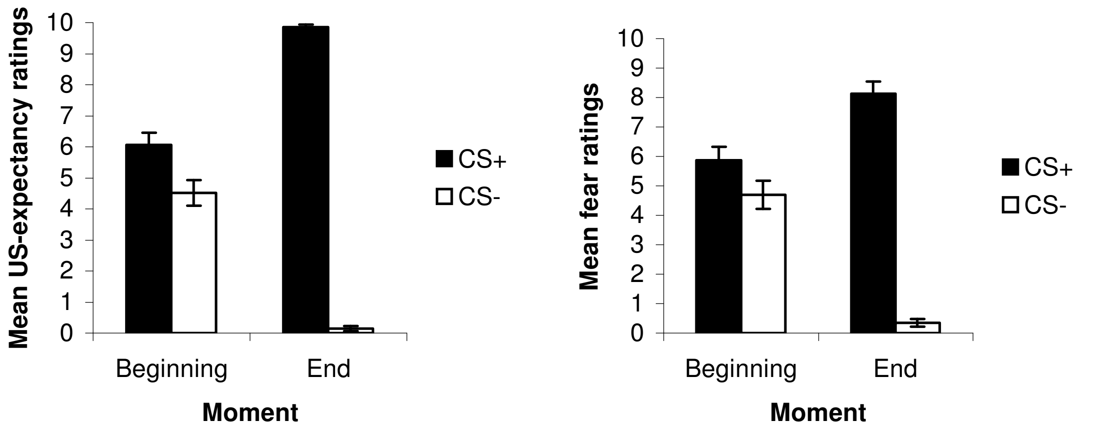


Figure 16. Mean US-expectancy (left panel) and fear (right panel) ratings for the CS+ and CS- at the beginning and end of acquisition. Error bars represent standard errors.

In line with expectations, participants' mean US-expectancy ratings for the beginning of the acquisition phase are close to the middle point of the scale, both for the CS+ and CS-. At the end of the acquisition phase, however, participants clearly expect the US to occur after the CS+, but not after the CS-. This pattern of results was confirmed by a 2 x 2 repeated measures ANOVA with moment (beginning/end) and CS-type (CS+/CS-) as within-subjects variables. The crucial Moment x CS-type interaction was highly significant, $F(1, 28) = 215.09$, $p < .0001$. Participants demonstrated significantly higher expectancy ratings for the CS+ as compared to the CS- both at the beginning, $F(1, 28) = 7.49$, $p < .05$, and at the end of acquisition, $F(1, 28) = 4315.26$, $p < .0001$, but this effect was far more pronounced at the end of acquisition. A strong increase in US-expectancy after the CS+ was observed from the beginning ($M = 6.07$, $SD = 2.12$) to the end ($M = 9.86$, $SD = .44$) of acquisition, $F(1, 28) = 107.42$, $p < .0001$. Participants' US-expectancy after the CS-, on the other hand, decreased sharply over time, $F(1, 28) = 101.90$, $p < .0001$, $M_{begin} = 4.52$, $SD_{begin} = 2.25$; $M_{end} = 0.14$, $SD_{end} = .44$. Finally, the analysis revealed a main effect of CS-type, $F(1, 28) = 337.62$, $p < .0001$. In general, the CS+ elicited more US-expectancy than did the CS-.

In sum, the retrospective US-expectancy ratings perfectly corroborate the findings of the online expectancy ratings.

Retrospective fear ratings. For the retrospective fear ratings, we expected to observe similar amounts of fear for the CS+ and CS- at the start of acquisition, and higher fear levels for the CS+ than for the CS- at the end of acquisition. As can be seen in Figure 16 (right panel), this predicted pattern was obtained in the data.

A 2 x 2 repeated measures ANOVA with moment (beginning/end) and CS-type (CS+/CS-) as within-subjects variables revealed a highly significant Moment x CS-type interaction, $F(1, 28) = 98.10$, $p < .0001$. Participants demonstrated higher fear ratings for the CS+ as compared to the CS- both at the beginning, $F(1, 28) = 4.98$, $p < .05$, and at the end of acquisition, $F(1, 28) = 331.51$, $p < .0001$, but this effect was far more pronounced at the end of acquisition. A strong increase in fear for the CS+ was observed from the beginning ($M = 5.86$, $SD = 2.50$) to the end ($M = 8.14$, $SD = 2.23$) of acquisition, $F(1, 28) = 16.19$, $p < .001$. Participants' fear for the CS-, on the other hand, decreased sharply over time, $F(1, 28) = 80.44$, $p < .0001$, $M_{begin} = 4.69$, $SD_{begin} = 2.56$; $M_{end} = 0.34$, $SD_{end} = 0.72$. Finally, the analysis revealed main effects of moment and CS-type, with $F(1, 28) = 6.46$, $p < .05$ and $F(1, 28) = 170.43$, $p < .0001$. In general, participants were more fearful at the beginning as compared to the end of acquisition and were more fearful for the CS+ than for the CS-.

Discussion

Results indicated that our RT task again succeeded in capturing the valence of the CSs during conditioning. In all three acquisition blocks, participants were found to respond faster to congruent probes than to incongruent probes. This finding was in line with participants' evaluative ratings, which indicated that the CS+ had acquired a negative connotation and the CS- a positive connotation by the end of conditioning. Already in the first acquisition block, participants were found to respond faster to congruent probes than to incongruent probes. There was a trend for an increase in the effect of affective congruence over blocks, but this trend was not significant. The latter findings suggest that participants quickly reached an asymptote in their learning but might also indicate that our RT task was perhaps not sensitive enough to capture small(er) changes in participants' evaluations over time.

Despite our efforts to draw participants' attention to the predictive value of the CSs, our RT task failed to capture participants' expectancies/fear online. Unlike in the original Dawson et al. (1982) study, participants were found to respond equally fast to CS+ and CS- probes. The online[48] and retrospective US-expectancy ratings clearly indicated that participants learned the differential CS-US contingencies. The retrospective fear ratings further suggested that participants became rather fearful of the CS+ (but not of the CS-) during conditioning. It is not clear why our RT task failed to capture this differential fear/expectancy *during* conditioning. A possible explanation might be that, despite our attempt to make the conditioning experience more primary and the RT task more secondary for the participants, the affective categorization task still reduced participants' cognitive engagement in anticipating the US deliverances during conditioning. The rating data indicate that participants became aware of the predictive CS-US relations, but this does not imply that participants' cognitive resources during the CS+ presentations were mainly devoted to predicting the US (which we assume to be the case in fear studies applying the original secondary probe technique or without secondary task).[49] We touch here upon a disadvantage of online measures in general: their assessment might interfere with the process under study. A similar concern, for instance, has been raised for the assessment of fear with the blink-startle measurement. Lipp (2006) pointed out that the loud acoustic probes that are used to elicit the startle reflex are often regarded as aversive by the participants and therefore might alter participants' experience during the experiment.

Our affective variant of Dawson et al.'s (1982) task might be more attention or resource demanding than the original probe technique for several reasons. One possible reason could be that a *categorization* task is more difficult than a simple detection task. Another explanation could relate to the *number of presented probes*. Dawson et al. (1982) used only a limited amount of probes (i.e., 16 probes in 40 trials) in their original study. Also in other studies

[48] Note that the online US-expectancy ratings were also retrospective (see procedure).
[49] The high fear ratings might seem to somewhat contradict this hypothesis. It seems logical to assume that highly fearful participants are likely to devote their (main) resources during conditioning to predicting/anticipating the feared event. Note that the validity of the fear ratings can be contested, however, as they were taken retrospectively.

that obtained good results with the secondary probe technique, probes appeared only occasionally (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004). In our procedure, we chose to work with a considerably larger number of probes (Experiment 2: 96 probes in 96 trials, Experiment 3: 60 probes in 36 trials) because we assumed that the collection of more RT data would lead to a more reliable valence/expectancy index. The use of more probes might, however, have increased the task's attentional and cognitive demands. Also the *affective nature* of our categorization task might have been responsible for the absence of an expectancy effect in our data. Some authors suggest that humans might be predisposed to primarily assign their attention to affective material as in daily life it has survival value to do so (Nairne & Pandeirada, 2008). These different factors might then have reduced participants' cognitive engagement in predicting/anticipating the US during the CS presentations, as a result of which the predictive value of the CSs failed to affect responding in our RT task (also see Spruyt, De Houwer, & Hermans, 2009). More research is needed to examine these hypotheses.

It is further important to note that even studies that more closely followed the original procedure of Dawson et al. (1982) have not always found consistent results with this task (e.g., Joos et al., 2008; Kerkhof, 2006). This suggests that the secondary probe technique of Dawson et al. (1982) might not be a very reliable task to measure participants' expectancies/fear or that its mechanism might depend on boundary conditions (e.g., number of probes, unpredictability of probes, modality of probes, timing of probes, difficulty of task) that are as yet poorly known. Therefore, to find out why we failed to obtain expectancy effects in our task and whether it is possible to create a secondary probe technique that can measure both participants' expectancies and evaluations, first more research is needed on the mechanism and properties of the original task.

Importantly, the findings of Experiment 2 and 3 clearly indicate that our affective secondary probe technique does constitute a valid *valence* measure that can be used to assess participants' evaluations online, during conditioning. In our view, this RT task forms a valuable addition to existing indirect attitude measures and may especially be a convenient tool for research that focuses on the formation and change of attitudes.

References

Dawson, M. E., Beers, J. R., Schell, A. M., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical-conditioning. *Journal of Experimental Psychology - General, 111*(3), 273-295.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11,* 549-554.

Fonteyne, R., Vervliet, B., Hermans, D., Baeyens, F., & Vansteenwegen, D. (2009). Reducing chronic anxiety by making the threatening event predictable: An experimental approach. *Behaviour Research and Therapy, 47,* 830-839.

Joos, E., Vansteenwegen, D., & Hermans, D. (2008, March). *Consolidation of the CS-US association through mental rehearsal*. Poster presented at the Associative Learning Symposium XII, Powys, Wales, UK.

Kerkhof, I. (2006). *Counterconditioning in a differential fear paradigm*. Unpublished data.

Kerkhof, I. (2008). *A secondary reaction time task to measure US-expectancy learning with visual probes*. Unpublished data.

Lipp, O. V. (2006). Human fear learning: Contemporary procedures and measurement. In M. G. Craske, D. Hermans & D. Vansteenwegen (Eds.), *Fear and learning: Basic science to clinical application* (pp. 37-52). Washington, DC: American Psychological Association.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Is survival processing special? *Journal of Memory and Language, 59*, 377-385.

Spruyt, A., Clarysse, J., Vansteenwegen, D., Baeyens, F., & Hermans, D. (2010). Affect 4.0: A free software package for implementing psychological and psychophysiological experiments. *Experimental Psychology, 57*, 36-45.

Spruyt, A., De Houwer, J., & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation *Journal of Memory and Language, 61,* 37-54.

4
General discussion

In this general discussion, the main aims and findings of the experimental studies conducted in both lines of research will be briefly recapitulated. We will highlight some potential theoretical, methodological, and clinical implications of our results as well as discuss limitations and directions for future research.

Changing conditioned preferences through counterconditioning

Background and research aims

In the General Introduction, we explained how in a fear conditioning procedure, the meaning of the CS alters in two important ways: the CS becomes a predictor for the aversive US, but also acquires a negative connotation through evaluative conditioning (EC). A child that was once bitten (US) by a dog (CS) may learn to expect other dogs to bite too and may acquire a dislike for dogs because of this experience. Acquired fears or phobias are generally treated using exposure therapy. In an exposure or extinction treatment, the CS is presented repeatedly in the absence of the US. The patient learns that the CS is no longer predictive of the US, which typically results in a gradual decrease of the fear responses. Remarkably, several studies indicate that the negative meaning that an object acquires in a fear conditioning procedure is little sensitive to such an exposure/extinction procedure. Hence, after exposure therapy, the CS may no longer be seen as a predictor for the US, but may still remain negative. The child from our example may no longer expect to be bitten every time he/she passes a dog, but may still dislike dogs. Importantly, some recent findings indicate that this negative valence that outlasts extinction/exposure might not be without consequences but might negatively impact the course and (short- and/or long-term) outcome of an extinction/exposure treatment. These findings then suggest that patients with a fear disorder might benefit from a treatment that not only focuses on the disconfirmation of expectancies through exposure, but also targets the acquired negative valence of the fear object.

A promising approach for changing the acquired negative valence of a phobic object seems to be counterconditioning (CC). In a CC procedure, the CS is paired with a US that has a valence opposite to that of the original acquisition US. In our example, CC could for instance entail presenting the child with his/her favorite type of candy every time he/she encounters a dog. A review of the literature on evaluative CC (see Chapter 1, Part 2) revealed, however, that the evidence for the effectiveness of this technique for changing valence is rather scarce and preliminary. Moreover, little is known on the sustainability (e.g., over time and contexts) of the possible evaluative effects that this procedure brings forth. Given the potential therapeutic merits of a technique that allows changing a person's existing (dis)likes, the first (and main) line of research of the present dissertation focused on a deeper investigation of the effects and properties of evaluative CC.

Four main goals were pursued. In a *first* step, we wanted to find a robust EC paradigm in which we could study the acquisition and change of conditioned (dis)likes. In a *second* step, we sought to examine whether we could replicate the finding that previously acquired conditioned valence can be altered by pairing the CS with a US that has a valence opposite to that of the original acquisition US. In a *third* step, we aimed at investigating the time- and context-sensitivity of preferences that were altered through CC. In a *fourth* and final step, we wanted to explore whether return of fear could effectively be reduced by presenting CC trials after extinction. In the next section, we summarize and discuss the main findings for each of these four research questions.

1) Finding a stable paradigm to study evaluative (counter)conditioning

The first two studies aimed at finding a *stable EC paradigm* that would allow studying evaluative CC.

In Experiment 1, we tested whether EC effects could be obtained using a picture-picture paradigm modeled after Walther, Gawronski, Blank, and Langer (2009). In this procedure, both the CSs and USs were pictures of human faces. Whereas the choice of the CSs was based upon ratings made by each individual participant, the USs were pre-selected. Before the start of the EC phase, the valence of the USs was boosted by pairing these US faces with several personality statements that were congruent with their a-priori valence. The goal of this US-formation phase was to create a strong and stable US valence. The US ratings indicated that this purpose was achieved: participants experienced the US faces as intended. Surprisingly, the CS ratings only revealed an EC effect for the CSs that were paired with a positive US. As expected, these CSs were evaluated more positively after conditioning (as compared to before conditioning). The evaluation of the CSs that were combined with a negative US did not change from pre-test to post-test, however. Furthermore, in the priming data, no significant EC effects were obtained at all. It is not clear why we failed to replicate the EC effects obtained by Walther et al. (2009). We hypothesized that a possible explanation might be found in two small alterations we made to their procedure: (1) we increased the number of acquisition trials and (2) used a delayed instead of a simultaneous conditioning procedure. More research is needed, however, to determine whether these factors might indeed have played a role. Because successful EC effects were obtained in the other two paradigms we tested (see Experiment 2), we did not further explore whether we could acquire stronger EC effects with the picture-picture procedure by making some alterations to it.

This experiment is not the first in failing to replicate EC effects; several failures to observe EC effects have been reported in the literature (e.g., Field & Davey, 1999; Field, Lascelles, Lester, Askew, & Davey, 2008; Rozin, Wrzesniewski, & Byrnes, 1998). Findings like these point out that the occurrence or strength of EC effects might depend on certain *boundary conditions* or *moderator variables*. In recent years, the focus in EC research has therefore shifted from the question of whether EC exists to the question of *when* an EC procedure will result in (weak or strong) changes in preferences (see Hofmann, De Houwer, Perugini, Baeyens, & Crombez, in press, for a meta-analysis). The findings of our first experiment make clear that a post-hoc identification of the factors that might have contributed to a replication failure is not easy. Seemingly small factors like the number of acquisition trials or certain temporal presentation parameters might be more important than appeared at first sight. Moreover, the presence of several contextual factors (e.g., attentional factors, contingency awareness, etc.) is difficult to compare between studies. It is clear that more controlled, comparative studies are needed to identify the possible importance of each of these variables. Hitherto, only a limited amount of studies have examined the role of certain procedural or contextual parameters in a *systematic* way (e.g., Lascelles & Davey, 2006). Like in our first study, in many other studies more than one variable is altered over experiments (e.g., Field et al., 2008), making it impossible to disentangle the effects of each individual factor. The present dissertation was not aimed at delineating boundary conditions or moderator variables in EC. Our findings do, however, provide several useful clues on what such variables might be (see above and further in this discussion) and therefore can serve as a source of inspiration for this type of research.

In a second experiment, we tested two gustatory EC paradigms. In the first, participants were asked to eat small cookies differing in color and shape during the conditioning phase. The color-shape compound of the cookies served as CS while the flavor of the cookies (good: honey or hazelnut; bad: Tween20 or vegetable bouillon) served as US. This procedure was used with success in previous studies conducted in our laboratory (e.g., Verhulst, Hermans, Baeyens, Spruyt, & Eelen, 2006) but has the disadvantage that the CS and US are difficult to disentangle because they form part of one and the same complex stimulus configuration. This paradigm is therefore less suitable for studying research questions that require separate CS or US presentations (e.g., the question of whether conditioned preferences can be altered through an extinction procedure). For this reason, we tested whether equally strong EC effects could be obtained with an adapted version of this paradigm in which the CS and US no longer formed a compound. In this second paradigm, the USs were still positively and negatively flavored food items, but the CSs were pictures of cookies instead of real color-shape compounds. The main findings of Experiment 2 indicated that differential EC was successfully demonstrated in both procedures. Both the rating and priming data confirmed that previously neutral CSs became significantly more liked if paired with a positive flavor and significantly more disliked if paired with a negative flavor. Because of the broader and easier applicability of the picture-flavor paradigm (i.e., the second paradigm), we chose to work with this procedure in the following studies (Experiments 3 and 4). Reliable EC effects were also found in those experiments, confirming the robustness of this paradigm.

Given that EC effects are not always easy to obtain, many authors care for a robust paradigm in which EC effects can be studied. Our findings indicate that the picture-flavor paradigm that was developed and applied in the present dissertation provides a promising tool for EC researchers.

Just as it is difficult to pinpoint the factors that might have been responsible for the absence of EC effects in a certain study, it is also not easy to identify the variables that might have contributed to positive effects. In the following paragraphs, we speculate on some characteristics of our picture-flavor paradigm that might have facilitated the occurrence of EC effects.

In our view, an important strength of our paradigm might lie in the *USs*. A prerequisite for EC effects to occur is that the USs are experienced as intended by the participants. In all of our studies, this precondition was met. The flavors moreover evoked strong (positive or negative depending on the exact flavor) reactions in the participants. Our centre has a long tradition of studying EC. A remarkable observation is that failures to obtain EC effects have mainly occurred in picture-picture, and in particular face-face studies (e.g., Experiment 1; Kerkhof, 2008). More stable results, on the other hand, were obtained when gustatory, olfactory or electrocutaneous US stimuli were used (e.g., Coppens et al., 2006; Hermans, Baeyens, Lamote, Spruyt, & Eelen, 2005; Vansteenwegen, Francken, Vervliet, Declercq, & Eelen, 2006; Verhulst et al., 2006). A possible explanation for this discrepancy in findings might be that participants experience the latter types of USs as more intense or salient than, for instance, a picture of a disliked face. In their meta-analysis of EC, Hofmann et al. (in press) examined the effect of US modality and only obtained evidence for stronger EC effects when an electrocutaneous stimulus was used as US. Comparable effects emerged for visual, gustatory and olfactory USs. However, as USs within one and the same modality may differ in intensity or salience (e.g., an unpleasant vs. an aversive flavor; a picture of a somewhat disliked face vs. of a mutilated body), Hofmann et al.'s analysis does not provide a strict test of the hypothesis that this variable might provide another moderator of EC effects and might have contributed to the strong and stable EC effects that were obtained in our picture-flavor studies.

In addition to US characteristics, certain *CS characteristics* or *features* of the *CS-US relation* might be crucial to obtain EC effects as well. In our picture-flavor paradigm, we paired pictures of differently shaped and colored cookies with a good or bad taste. Even though Baeyens, Eelen, Van den Bergh, and Crombez (1990) applied similar USs (e.g., Tween20 and sugar drinks), they failed to obtain EC effects when colors were used as a CSs. Further research is needed to investigate whether this difference in results might relate to certain characteristics of the applied CSs or of the CS-US relation. Possible factors that may play a role are, for instance, CS familiarity and/or CS-US belongingness. Some studies suggest that EC effects are more difficult to obtain for familiar CSs than for unfamiliar CSs (e.g., Cacioppo, Marshall-Goodell, Tassinary, & Petty, 1992; Stuart, Shimp, & Engle, 1987). This might explain why the evaluation of colors is more difficult to change through EC than the evaluation of pictures of unfamiliar cookies. There is also some evidence that CS-US belongingness might moderate EC effects. Todrank, Byrnes, Wrzesniewski, and Rozin (1995), for instance, used photographs of human faces as CSs and odors as USs and found evaluative shifts to occur only when the odors used were ‘plausibly human’ (e.g., chemical imitations of naturally produced odors such as sweat or scented products applied to the body such as soap). No EC effects were obtained when odors typically associated with objects rather than with people were used. Similarly, Lascelles, Field, and Davey (2003) found evaluations of food pictures to change only when they were paired with pictures of obese body shapes but not when combined with thin body shapes. Belongingness or the presence of some sort of ‘conceptual connection’ between the CS and US might thus facilitate EC effects. The CSs (pictures of cookies) and USs (real cookies) in our picture-flavor paradigm are closely connected. Hence, also this factor might constitute a strength of our paradigm and explain the discrepancy with the results of Baeyens et al. (1990), where the connection between the CS and US was less obvious.

Finally, another factor that might have played a crucial role is *contingency awareness*. Our contingency recall data suggest that most likely the majority of the participants were aware of the CS-US contingencies during learning. The recall data indicated that participants remembered most of the experimental contingencies. Moreover, since we used a post-conditioning contingency recall assessment, participants’ level of contingency awareness during the experiment might even be underestimated due to forgetting.[50] Several authors argue that contingency awareness is necessary for EC effects to occur (e.g., Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Wardle, Mitchell, & Lovibond, 2007; but see Fulcher & Hammerl, 2001; Walther & Nagengast, 2006 for a different opinion). Supporting this idea, the meta-analysis of Hofmann et al. (in press) identified contingency awareness as the most important moderator of EC effects. Therefore, the fact that participants could easily distinguish the experimental contingencies in our experiments might also have contributed to the strong EC effects that were obtained. Note that this latter finding also entails the danger that demand effects might have been at play in our rating results. In this context it is important to note that the rating results were always backed up with priming data, which are much less prone to demand bias.

It is clear that the above-mentioned considerations are only speculative and require further empirical investigation.

2) Changing conditioned preferences through counterconditioning

As mentioned above, relatively few studies have examined the impact of CC on evaluative learning in human beings. The studies that were conducted suggest that CC provides a promising technique to alter previously conditioned valence (e.g., Baeyens, Eelen, Van den Bergh, & Crombez, 1989; Lipp & Purkis, 2006; but see Stevenson, Boakes, & Wilson, 2000 for a failure to change conditioned preferences through CC). In Experiment 3, we examined whether we could *replicate* this finding. Using the picture-flavor paradigm of Experiment 2, we compared the effect of further conditioning, extinction and CC after a

[50] Note that we are aware of the fact that one should be careful in using contingency recall data as an index of contingency awareness. As mentioned, because of forgetting, retrospective memory tests may not provide an *exhaustive* measure of the contingency awareness that was available during learning (e.g., Field, 2000; Shanks & St. John, 1994). They may also not provide an *exclusive* measure of contingency awareness. A memory test might be biased by guessing or participants may retrospectively construct assumptions about the contingency on the basis of, for example, their attitude toward the CS rather than on actual knowledge about the pairing during the learning task (e.g., Bar-Anan, De Houwer, & Nosek, 2009). For this reason, we also refrained from using our contingency recall data to examine and formulate conclusions about the role of contingency awareness in EC (and CC).

differential[51] EC acquisition phase. Results indicated that whereas further conditioning and extinction trials failed to eliminate the valence the CSs acquired during acquisition, the CC treatment did. The latter treatment even succeeded in reversing the CSs' initially acquired valence. These effects appeared both in the CS ratings and in an affective priming task (APT). To our knowledge, this is the first study to demonstrate evaluative CC effects both in a direct and indirect valence measure. Our findings thus corroborate and extend previous research that found EC to be sensitive to CC. Note that also in our fourth experiment (discussed in more detail in the following section) we found CC to be effective in altering previously acquired conditioned valence, confirming the reliability of this finding.

As indicated in the above paragraph, in Experiment 3 we found EC to be little sensitive to extinction. This finding is in line with several previous studies (e.g., Diaz, Ruiz, & Baeyens, 2005; Vansteenwegen et al., 2006; for other examples see General Introduction). Nevertheless, it remains a remarkable observation as in other forms of Pavlovian conditioning repeated CS-alone presentations after acquisition are known to result in a strong decrease in conditioned responding (e.g., Hamm & Vaitl, 1996). Later in this discussion, we will return to this issue in more detail.

3) The sustainability of counterconditioning effects over time and contexts

The few previous studies that examined CC in EC limited themselves to a (successful or not so successful) demonstration of the phenomenon (e.g., Baeyens et al., 1989, Lipp & Purkis, 2006; Stevenson et al., 2000). Hitherto, no research has been conducted on the *durability* of the evaluative effects that this procedure brings forth. In our third and fourth experiment, we addressed this gap in the literature and investigated the stability of CC effects over time and contexts.

In Experiment 3, we examined the *time-sensitivity* of counterconditioned preferences by assessing valence at 1-week follow-up and found the effects of the CC treatment to persist over this time period. There was no evidence for the

[51] Half of the CSs were paired with a positive US while the other half were combined with a negative US.

spontaneous recovery[52] of participants' initially (during the acquisition phase) acquired evaluations. This finding was evidenced by both evaluative rating data and priming data.

In Experiment 4, we investigated the *context-sensitivity* of CC effects in our picture-flavor paradigm. Two groups were compared: (1) an AAA group that received differential acquisition training, CC and test in a single context A and (2) an ABA group that underwent acquisition training in context A, CC in context B and that subsequently returned to context A for test. Different contexts were created by the display of images of different bakeries on the background screen. For half of the participants, evaluative ratings were collected at the start, middle and end of each phase. The other half of the participants was required to give US-expectancy ratings at these moments. In previous studies, ABA renewal effects (after extinction *and* CC) have been observed for expectancy learning (e.g., Peck & Bouton, 1990; Vansteenwegen et al., 2005). The expectancy group was therefore included as a manipulation check to verify that renewal effects could be obtained in our paradigm. At the end of the experiment, participants' evaluations were also assessed indirectly in both contexts with an APT.

The rating results indicated that the CC treatment succeeded in changing participants' initially acquired expectations and preferences. These data further revealed an ABA renewal effect in the US-expectancy ratings but not in the valence ratings. Upon return to the original acquisition context, participants' initially acquired expectations reappeared in the ABA (but not AAA) group. Participants' evaluative ratings, on the other hand, proved to be insensitive to the context switch and were still in line with the contingencies of the CC phase. The evaluative rating data thus suggest that counterconditioned preferences can survive a context switch. Also in the priming data, no evidence was obtained for contextual modulation of evaluative learning. The findings of the priming task should be interpreted with caution, however, as they only revealed null effects. Therefore, we cannot exclude the possibility that our priming task simply failed to pick up participants' evaluations. The bakery background pictures might, for instance, have interfered with the priming task by distracting participants'

[52] Note that the terms 'spontaneous recovery' and 'renewal' are again being used here to refer to the possible reappearance of an initially acquired conditioned response after *counterconditioning* (not *extinction*).

attention from the prime stimuli or categorization task. Further research (e.g., using more simple background pictures) is needed to clarify this issue.

In sum, our third and fourth experiment found the evaluative effects of a CC treatment to remain intact over time and after a context switch. Importantly, these findings suggest that CC might induce a rather stable and durable change in conditioned valence. As this is the first work to examine the sustainability of CC effects in EC, further research will be needed to confirm the present findings. A limitation of Experiment 4, for instance, is that, due to the fact that our priming data were difficult to interpret, our conclusion that the evaluative effects of a CC treatment can survive a context switch is mainly based on self-report data. This leaves open the possibility that our results in that study could have been driven by demand effects (but see Chapter 4 of Part 2 for arguments against this vision) and supports the idea that replication studies are needed.

In the General Introduction, we noted that the findings on the time- and context-sensitivity of CC in EC might provide some indication on how second-learned evaluations are represented in memory. We will address this topic later on in a separate section of this discussion. In that section we will also take a closer look at the remarkable finding in our fourth experiment that ABA-participants' *expectancy* and *evaluative* ratings demonstrated a different pattern of results upon return to the A context.

4) Reducing return of fear through counterconditioning

A final goal in this first line of research was to explore whether we could obtain evidence for our hypothesis that *return of fear might be reduced* by targeting the CS's extinction-resistant negative valence with a CC procedure. In a fifth and last experiment, we examined this question in a reinstatement study in mice. After a differential fear acquisition and extinction procedure with a foot shock as US, half of the animals (CC group) received CC trials during which the CS+ was paired with food pellets, while the other half (FE group) underwent further extinction trials. Subsequently, half of the mice in each treatment group (CC-reinstatement and FE-reinstatement groups) received unsignaled USs while the other half (CC-control and FE-control groups) did not. Based on the higher described framework, we predicted the CC trials to eliminate the acquired negative valence of the CS+ and therefore expected to observe less return of fear for this stimulus in the CC-reinstatement group as compared to the FE-reinstatement group. In contrast to this prediction, however, no beneficial effect was found of the CC treatment on return of fear due to reinstatement. An equally strong return of conditioned fear responding was observed in both reinstatement groups. This finding could be seen as evidence against our hypothesis that the remaining negative valence of the CS+ after extinction might form a source for return of fear. However, as CS valence was not measured in this experiment, we cannot exclude the possibility that the CC trials perhaps failed to (sufficiently) alter the negative valence that the CS+ acquired through its pairings with the foot shock. More CC trials or a stronger positive US might be required to obtain a change in valence that results from pairings with such a strong biologically significant US. Future research should include a CC manipulation check to tease apart these two different explanations.

Clearly, it would also be interesting to investigate the possible impact of changing a CS's negative valence through CC on return of fear in humans. There is actually one (quasi-)experimental study that addressed this question. De Jong, Vorage and van den Hout (2000) compared a group of spider phobics that received a standard one-session exposure treatment to a group that received an equally long treatment that consisted of exposure and CC. In the CC group, tasty food-items were used during the standard exposure exercises and the participants' favorite music was played. In contrast with the predictions of our framework, no difference in treatment efficacy was observed between both groups when tested immediately after treatment and at 1 year follow-up. Moreover, CC was not found to be more effective in altering the affective valence of spiders than regular exposure treatment; both procedures resulted in a significant reduction in the negative evaluation of spiders. There are, however, some important limitations to the De Jong et al. (2000) study. A first limitation is that the standard exposure sessions probably included several ingredients that may have helped to undermine the negative affective valence of spiders and can be conceptualized as forms of CC (e.g., accepting expressions by the therapist towards spiders, information about spiders that portrays them as tender, fragile and timid animals). A second remark is that both procedures altered the affective evaluation of spiders in a positive direction, but did not lead to the abolishment of spiders' negative valence; an important amount of negative valence remained after both treatments. Perhaps the CC procedure was too brief in duration or not of sufficient strength to neutralize the spiders' negative valence. Also here, further research is needed to evaluate and clarify the (potential) role of stimulus valence in return of fear.

Both the findings of our mice study and those of De Jong et al. (2000) suggest that an interesting avenue for future research might also lie in examining how the evaluative effects of a CC procedure can be boosted (e.g., perhaps by increasing the number of CC trials, by pairing the CS with multiple positive USs, etc.).

What is learned in an evaluative (counter)conditioning procedure?

In the General Introduction, we explained that a CC procedure can be conceptualized as a retroactive interference paradigm in which participants learn conflicting associations in two temporally separated phases. Studies on Pavlovian conditioning typically show that in such interference paradigms (like CC but also extinction), the second learning phase does not involve 'unlearning' but rather entails the learning of a new conditioned response that temporarily suppresses the former one. Evidence for this comes from studies on rapid reacquisition, spontaneous recovery, renewal, and reinstatement, phenomena that illustrate that the original conditioned response can quickly be restored with new learning trials or can suddenly re-emerge after the mere passage of time, a context change, or unsignaled presentations of the original US (for a review, see Bouton, 2002, 2004, all these phenomena have been observed after extinction as well as after CC). According to modern learning psychologists (e.g., Bouton, 2002), interference paradigms result in the creation of an 'ambiguous' CS. Temporal or physical context elements will then determine which association (i.e., the first- or second-learned) is retrieved and guides responding. Pavlovian conditioning studies further suggest that in such interference paradigms, second-learned information is more time- and context-dependent than first-learned information. According to Bouton (2002, 2004), our learning and memory system treats the first-learned information (e.g., CS-US or CS-US_1) as context-free, but the second-learned information (e.g., CS-noUS or CS-US_2) as a kind of time- and context-specific 'exception to the rule'. The data of the ABA-*expectancy* group in Experiment 4 are in line with this account. The CC treatment in context B succeeded in changing participants' initially (in context A) acquired expectations. The latter expectations reappeared, however, upon return to the original acquisition context. These data support the idea that participants' first-learned conditioned expectations were not destroyed but only temporarily suppressed by the CC trials in context B.

Remarkably, a different pattern of results was observed for participants' conditioned *evaluations* in Experiment 4. Also here, the CC trials in the B context were found to be effective in altering participants' initially learned evaluations. No evidence was found, however, for a return of these first-learned evaluative responses upon re-testing in the A context (put differently: no renewal effect occurred). A similar surprising result was obtained in Experiment 3, where we failed to observe a spontaneous recovery effect in our valence indices when participants' evaluations were assessed 1-week after the CC treatment. Instead, the evaluative effect of the CC procedure was found to remain intact over this time period. Importantly, both findings suggest that second-learned evaluations might be less time- and context-sensitive than other second-learned conditioned responses (e.g., preparatory responses) and perhaps even able to permanently overwrite/replace the first-learned evaluation. As we were the first to investigate whether ABA renewal and spontaneous recovery effects could be observed after CC in EC, further research is needed to confirm these findings and to allow definite conclusions about what is learned in the CC phase of an EC procedure.

Interestingly, our findings seem to suggest that EC might be governed by different learning principles than other forms of Pavlovian conditioning and therefore support the claim of several authors that EC constitutes a unique form of associative learning (e.g., Baeyens & De Houwer, 1995; Martin & Levey, 1994). The fact that we found EC effects to be little affected by extinction trials in our third experiment[53] lends further weight to this claim. Important to note, however, is that there are also several findings in the literature that contest the

[53] Note that in Experiment 2 of our second line of research we also failed to find an extinction effect in our online indirect valence measure.

idea that EC would be different from other types of Pavlovian conditioning. For instance, in contrast with the findings of our third and fourth experiment, Lipp and colleagues (Hardwick & Lipp, 2000; Lipp, Oughton, & Lelievre, 2003) did obtain evidence for extinction and modulation effects in EC. In the General Introduction, we already pointed out that the literature is fraught with inconsistent findings concerning the functional characteristics of evaluative associative learning (e.g., (in)dependent of contingency awareness, resistant or susceptible to extinction, modulation, cue competition, etc.). Hitherto, it is not yet clear how to reconcile this wide range of conflicting results. Some authors have suggested that methodological or procedural factors might contribute to the discrepant findings (e.g., Lipp et al., 2003). The differences in results between our studies and those of Lipp and co. might, for instance, relate to the different dependent variables that were used to assess participants' evaluations.[54] Hardwick and Lipp (2000), for example, applied the startle response as dependent variable. As discussed in Part 3 of this dissertation, some authors have argued that this response does not provide a good index of EC because it can be affected by factors other than stimulus valence (e.g., arousal, Cuthbert, Bradley, & Lang, 1996). The use of this particular measure might then explain why Hardwick and Lipp (2000) did observe modulation of evaluative learning, where we failed to find this in Experiment 4. Another example can be illustrated with the study of Lipp et al. (2003). In the latter study, Lipp et al. obtained evidence for extinction effects in EC using online evaluative ratings as dependent variable. In Experiment 2 of our second line of research, on the other hand, we found EC to be resistant to extinction using an online indirect valence measure. A possible explanation for this discrepancy in results might be that demand effects biased the results of Lipp et al.'s (2003) direct measure or that our indirect reaction time task was not sensitive enough to pick up an extinction effect. An important area for future research lies in examining whether such procedural/methodological factors might indeed explain (some of) the observed differences in the functional

[54] Many more other procedural/methodological factors apart from the applied dependent variable might have been responsible for the difference in results (e.g., US modality, number of trials, type of instructions,…). Again, a post-hoc identification of these variables is not easy as EC studies generally differ from each other in several procedural/methodological aspects. Systematic research is then needed to single out the possible effect of each of these variables.

properties of EC effects. It is clear that EC research will not only benefit from studies that focus on identifying variables that moderate the strength or occurrence of EC effects (as was argued earlier in this discussion), but also from studies that search for variables that might moderate the functional properties of EC (see also De Houwer, 2007).

Some authors have argued that the opposing results in the EC literature might also reflect the operation of different underlying processes.[55] De Houwer and colleagues (De Houwer, Baeyens, & Field, 2005; De Houwer, 2007, 2009), for instance, suggest that in some studies EC effects might depend on the automatic formation of associations while in other studies the effects might be based on the formation of conscious propositional knowledge about the CS-US relation(s) (or even other processes). An explanation for the difference between our findings and those of Lipp et al. (Hardwick & Lipp, 2000; Lipp et al., 2003) might then, for instance, be that extinction and modulation occurs only for EC effects that rely on the formation of conscious propositional knowledge about the (conditional) CS-US relation. In our studies, on the other hand, EC effects might have been driven by a more simple learning process (e.g., the automatic formation of associations governed by a rudimentary Hebbian learning rule, see also Baeyens, Hendrickx, Crombez, & Hermans, 1998) that is not sensitive to manipulations that affect the statistical contingency between the CS and US. Note, however, that this explanation is post-hoc and that there are equally good reasons to assume that there was room for propositional reasoning/learning in our studies. The post-conditioning recall data of our different EC studies, for example, suggest that the vast majority of the participants were aware of the CS-US contingencies in our study (which is a prerequisite for propositional reasoning being able to drive EC effects). The fact that this precondition was met does not imply, however, that participants' preferences were necessarily rule-based. Although many problems are associated with (retrospective) verbal reports, a post-experimental questionnaire of participants' thoughts and reasons behind the evaluations that they reported in our experiments might have provided some more insight into this issue (and might thus be worthwhile to

[55] Note that this explanation is not incompatible with the previous one. Which processes are involved in an EC study might depend on certain methodological/procedural factors.

include in future research). It is clear, however, that determining what process underlies particular EC effects is not an easy task. The fact that EC theories are often ill-specified and difficult to derive predictions from, adds to the difficulty (also see Hofmann et al., in press). The propositional account, for instance, is relatively mute concerning the exact nature of the propositions that are assumed to drive EC effects. In early versions of this account, for instance, it has been argued that propositions about the statistical contingency between the CS and US are likely to mediate EC effects (e.g., De Houwer et al., 2005). More recent formulations of this theory, however, suggest that the propositions that underlie EC might also be limited to the fact that the CS and US co-occur (e.g., De Houwer, in press). In the former case, one would predict propositionally based EC effects to be sensitive to manipulations that affect the statistical contingency between the CS and US, in the latter situation not. The other EC theories suffer from a similar lack of precision/clarity (see Hofmann et al., in press).

In sum, the aforementioned considerations make clear that research on EC is actually still in its early stages and that much more empirical and theoretical work will be needed to fully understand the associative route of preference acquisition and change.

Some practical and clinical implications

As mentioned in the General Introduction, despite the limited research on the evaluative effects of a CC treatment, this technique is already frequently used in applied settings with the aim of changing people's existing evaluations. A common advertising strategy, for example, is to pair a product or corporation that has acquired a negative connotation with positive messages/images in commercials or campaigns. Also in clinical practice, CC techniques are regularly employed (e.g., Korrelboom, van der Gaag, Hendriks, Huijbrechts, & Berretty, 2008). Importantly, our findings support the validity of this strategy for altering a person's previously acquired conditioned (dis)likes. Moreover, our data suggest that CC might induce a rather durable change in valence that is maintained over time and generalizes well to other contexts. For clinical practice (but also other applied settings), this finding is reassuring as it suggests that new evaluations that patients acquire through CC can outlast the therapeutic context.

In two of our experiments (Experiment 3, line 1; Experiment 2, line 2) we found conditioned valence to be little susceptible to extinction. These observations reinforce the idea that the negative acquired valence of a phobic object can remain intact after a standard exposure treatment. Our findings are less clear, however, concerning the possible role of this remaining negative valence in return of fear. In our mice study, we found no beneficial effect of CC trials after extinction on return of fear. Nevertheless, due to the lack of a manipulation check, we cannot exclude the possibility that our CC manipulation failed to alter the negative valence of the CS+ in that study. Interestingly, this finding suggests that it might be more difficult to alter conditioned preferences that were acquired through pairings with a strong negative US (e.g., an electrocutaneous stimulus, a traumatic experience). As was mentioned before, an interesting path for future research might then lie in examining how the evaluative effects of a CC procedure can be boosted. The question of whether return of fear can be reduced by targeting the negative valence of the phobic object after extinction thus remains largely open. More experimental and clinical research (in animals *and* in human beings) is needed to verify the correctness of this hypothesis. The experiments that were reported in the present dissertation facilitate this type of research by providing a procedure that allows changing the negative valence of a fear stimulus.

From a clinical perspective, it might also be interesting to examine the role of stimulus valence in fear learning in more general. In the introductory chapter, we hypothesized that the negative affective valence of a fear stimulus might not only impact the long-term outcome of an exposure treatment (i.e., return of fear), but might also influence the course of this treatment. Hitherto, no research has been conducted, however, on the possible influence of stimulus valence on *extinction learning*. Similarly, it seems plausible that an object's evaluative connotation might also impact *fear acquisition*. A negatively valenced stimulus might, for instance, be more vulnerable for aversive conditioning while a positive valence may protect an object from becoming a fear cue. These ideas are speculative but, in our opinion, intriguing for further research as they may provide more insight into fear mechanisms.

Assessing valence indirectly and online

Background and research aims

In a second line of research, we focused on the assessment of valence. Indirect reaction time (RT) measures like the APT and IAT are very popular in evaluation research due to their easy application and the fact that they are assumed to be less vulnerable to demand effects than rating scales. In EC studies, these measures are typically administered as pre-test and/or post-test in a session that precedes or follows the evaluative learning phase. In the third part of this dissertation we discussed several disadvantages of such pre-test/post-test designs (e.g., these studies provide little insight into the course of evaluative learning, repeated administrations of indirect RT tasks are known to result in smaller effects, context effects might bias the measurement results, etc.). Based on these drawbacks, we argued that it sometimes might be preferable to assess valence online, *during* the learning phase. The currently existing indirect RT tasks, however, do not lend themselves well to be integrated in an ongoing EC procedure. Therefore, our second line of research was aimed at developing an indirect RT task that *can* be used to assess valence online.

Assessing valence indirectly and online

In our search to develop an online indirect valence measure, we were inspired by the work of Dawson, Beers, Schell, and Kelly (1982). These authors measured the operation of cognitive processes in a differential fear conditioning procedure by asking participants to respond to tone probes that were presented during the CSs. We reasoned that an affective variant of this measure could be created by replacing the neutral tone probes with affect-laden (i.e., positive and negative) visual probes. The in this way obtained task can be conceptualized as an online variant of the APT with the CSs of the conditioning procedure serving as primes and the visual probes as targets. The main difference with a ‘classic’ APT is that in our task the prime is not presented shortly before the target stimulus, but functions as a background stimulus against which the target stimuli appear. Like in a regular APT, we predicted priming effects to occur in our task and expected participants to respond faster to probes that were evaluatively congruent with the CS than to evaluatively incongruent probes.

In a first pilot study, we tested the validity of our developed task by examining its sensitivity to the evaluation of items that are normatively regarded as favorable or unfavorable (i.e., IAPS pictures; Lang, Bradley, & Cuthbert, 2005). Results indicated that our indirect RT measure was successful in grasping the valence of the IAPS pictures. As expected, participants responded faster when prime and probe were affectively congruent as compared to when they were incongruent. Based on these results, we can conclude that our task can be used to assess stable *pre-existing* attitudes.

In a second experiment, we incorporated our indirect RT measure in an EC procedure to test whether it was able to assess participants' attitudes *online*, *during* evaluative learning (the purpose this task was developed for). In a differential fear conditioning procedure with an electrocutaneous stimulus as US, participants were asked to categorize positive and negative icons that appeared at different timings during the CSs. In line with the results of several other studies, we expected the reinforced CS (CS+) to acquire a negative connotation because of its association with the aversive US, and the unreinforced CS (CS-) to remain neutral or become slightly positive (e.g., Hermans, Spruyt, & Eelen, 2003; Vansteenwegen et al., 2006). Because of the controversy surrounding the extinction resistance of EC (see discussion line 1), we also added an extinction phase to the conditioning procedure in which both CSs were presented unreinforced. Results indicated that our online RT task succeeded in tracking the expected changes in valence of the CSs. The RT data of the acquisition phase evidenced a gradual shift in the affective meaning of the CSs that resulted in faster responses to congruent probes[56] than to incongruent probes in the second part of acquisition. This effect was maintained during both halves of the extinction phase, which suggests that the previously acquired valence of the CSs outlasted extinction. This finding corroborates the results of Experiment 3 in our first line of research and those of several other studies that found no impact of an extinction procedure on EC effects (e.g., Diaz et al., 2005; Vansteenwegen et al., 2006). The possible theoretical meaning of the latter finding was already discussed above (again, see discussion line 1). Most importantly, we can

[56] Congruent probes = negative probes for the CS+ and positive probes for the CS-; Incongruent probes = positive probes for the CS+ and negative probes for the CS-.

conclude from this study that we succeeded in creating an indirect valence measure that can be used to assess changes in valence online. The latter conclusion was confirmed in a third study in which we again administered our task during a differential fear acquisition procedure. Also in this experiment, faster responses were observed to congruent probes than to incongruent probes in all three blocks of the acquisition phase.

In our opinion, our affective variant of Dawson et al.'s (1982) secondary probe technique forms a valuable addition to existing attitude measures and may especially be a convenient tool for research that focuses on the formation and change of attitudes. It allows for an indirect measurement of participants' evaluations and overcomes the afore-described problems (e.g., context effects) that are associated with attitude measures that can only be applied in a pre-test post-test manner. Other advantages are the fact that our RT task is easy in use and has the potential of providing more insight into the course of evaluative learning.

Further research should focus on ways to improve and learn more about our developed task. In the previous chapters we already discussed that it is remarkable that priming effects were observed in our RT procedure, given the long SOAs (Stimulus Onset Asynchrony, the time interval between the beginning of the prime and the target) that are applied in it. Due to the fact that the probes (the targets) were presented at different timings *during* the CSs (the primes) SOAs varied between 300 and 6000 ms in our different experiments. In this context, it is important to note that in priming studies the SOA is generally kept very short to minimize the potential for controlled response strategies to influence the RT data. A SOA of 300 ms or less is assumed to be too brief for subjects to have the time to intentionally process and strategically use the valence of the prime (Neely, 1977; Posner & Snyder, 1975). This argument is therefore often used to argue that priming effects obtained at a short SOA are unlikely to be biased by demand effects. As mentioned earlier, SOA length in our task was considerable longer than 300 ms. An important question is then whether this makes our RT task more vulnerable to demand effects. It would be interesting for future research to examine whether participants are aware of what our task is measuring and whether they are able to strategically influence the (size of the) priming effects in it.

Another potential limitation of our online measure (that will apply to any online RT task) is that RTs need to be aggregated across trials to obtain a stable valence index. This renders a fine-grained trial-by-trial view on the time course of evaluative learning impossible. In addition, the number of data points that can be collected to index valence will always be limited and dependent upon the number of learning trials. This might make our RT task vulnerable to sensitivity and reliability problems, which are already known to plague priming (but also other indirect attitude) measures (e.g., Cunningham, Preacher, & Banaji, 2001). Given that we found no influence of probe timing or number in our results, a good approach for further applications of our measure might be to present double (or perhaps even more) probes on all trials (but see next section for a possible disadvantage of this suggestion). Another avenue for future research could be to focus on creating an online version of other behavioral valence measures that tend to show better sensitivity/reliability than priming tasks (e.g., the IAT or the affect misattribution procedure of Payne, Cheng, Govorun, & Stewart, 2005).

Last, but not least, further research could investigate the application of our online measure in other paradigms that are used to induce changes in valence. It would, for instance, be interesting to examine whether and how our task can be integrated in the picture-flavor paradigm that we developed in our first line of research (and to see if we can replicate our CC results with it).

Assessing valence and expectancy indirectly and online

As mentioned earlier, Dawson et al. (1982) originally constructed their secondary task technique to assess the operation of cognitive processes during human classical conditioning. They presented tone probes during the CSs of a differential fear conditioning procedure and found participants to respond slower to probes presented during the reinforced CS than to probes presented during the unreinforced CS, indicating larger resource allocation during the CS+. In recent fear research, this task is commonly used as an indirect index of expectancy learning (e.g., Dirikx, Hermans, Vansteenwegen, Baeyens, & Eelen, 2004; Lipp, Siddle, & Dall, 1993). The affective version we developed from Dawson et al.'s probe technique is procedurally still very similar to the original task, the main difference being that the neutral probes were replaced by affect-laden ones. Therefore, we assumed our adapted task to still be sensitive to expectancy learning.

In the third chapter of Part 3, we argued that much benefit could be obtained from an indirect RT measure that is able to assess *both* evaluative learning and expectancy learning. Hitherto, these two types of learning have typically been studied in very dissimilar research designs using very different dependent variables. Some authors have argued that the observed differences between both types of learning might therefore be due to methodological or procedural factors rather than to true process differences (e.g., Lipp & Purkis, 2005). To minimize the potential for such factors to bias the results, it is advisable to assess evaluative and expectancy learning within a single design, using comparable dependent variables. Finding comparable measures for these both types of learning is, however, not obvious. We reasoned that if our affective version of Dawson et al.'s (1982) task would still represent a valid measure of expectancy learning, it would provide a promising tool for comparing evaluative learning and expectancy learning in a fair manner. The task can easily be integrated in an ongoing conditioning procedure and would allow the assessment of both types of learning in the same response domain.

We reanalyzed the data of Experiment 2 (the first of our two differential fear conditioning studies, see previous section) to see if participants indeed showed slower responses to probes presented during the reinforced CS as compared to probes presented during the CS-. Results indicated, however, that participants responded equally fast to CS+ and CS- probes. Hence, we failed to observe the standard expectancy effect that was repeatedly found in studies that applied the original secondary task technique of Dawson et al. (1982).

We hypothesized that a possible reason for the absence of an expectancy effect in our RT data might have been that the affective categorization task was no longer secondary in our study and perhaps captured too much of participants' resources during conditioning. In other words, our RT task might have prevented participants from adopting a 'predictive' mindset with a focus on predicting/anticipating the US during the CS presentations. To test this hypothesis, we conducted a third study (also see above) in which we replicated the design of Experiment 2, but made some crucial methodological changes that aimed at making the predictive component of the conditioning procedure more salient. For example, the primacy of paying attention to the CS-US relations was emphasized in the instructions, participants were required to fill in trial-by-trial US-expectancy ratings and participants were encouraged to select a strongly aversive US. We again explored whether our task was able to track both participants' evaluations and expectancies during conditioning. As already mentioned in the previous section, the RT data indicated that our online measure succeeded in capturing the valence of the CSs. However, despite our efforts to draw participants' attention to the predictive component of the conditioning procedure, our task once more failed to capture participants' expectancies online. As in Experiment 2, participants were found to respond equally fast to CS+ and CS- probes.

It is not really clear why we again failed to obtain an expectancy effect in our data. We argued that a possible explanation might be that, despite our attempt to make the categorization task more secondary in our third experiment, it still reduced participants' cognitive engagement in anticipating the US deliverances during conditioning. We discussed several reasons as to why our affective probe task might be more attention or resource demanding than Dawson et al.'s (1982) original task. One possible reason could, for instance, be that a *categorization* task is more difficult than a simple detection task. Another explanation could relate to the number of presented probes. In our experiments we chose to work with a considerably larger *number of probes* than in the original Dawson et al. (1982) procedure because we assumed that the collection of more data points would result in a more reliable valence/expectancy index. The use of more probes might, however, also have increased the attentional and cognitive demands of the task. Also the *affective nature* of our categorization task might have been responsible for the absence of expectancy effects. Some authors have suggested that humans might be predisposed to principally assign their attention to the affective dimension of a stimulus as in daily life it has survival value to do so (Nairne & Pandeirada, 2008). These different factors might then have reduced participants' cognitive engagement in predicting/anticipating the US during the CS presentations, as a result of which the predictive value of the CSs failed to affect responding in our RT task (also see Spruyt, De Houwer, & Hermans, 2009). Further research is needed to examine these hypotheses.

Important to note is that several studies that more closely followed the original procedure of Dawson et al. (1982) also failed to observe slower responses to CS+ probes than to CS- probes (e.g., Kerkhof, 2006). This suggests that Dawson et al.'s secondary probe technique might not be a very reliable task or that its mechanism might depend on certain boundary conditions (e.g., number of probes, unpredictability of probes, modality of probes, timing of probes, task difficulty) that are as yet poorly known. To find out why we failed to obtain an expectancy effect in our task and whether it is possible to create a secondary probe technique that can be used to measure both participants' expectancies and evaluations, future research should first focus on learning more about the mechanism and properties of the original task.

References

Baeyens, F., & De Houwer, J. (1995). Evaluative conditioning is a qualitatively distinct form of classical-conditioning - Reply. *Behaviour Research and Therapy, 33*(7), 825-831.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1989). Acquired affective evaluative value - Conservative but not unchangeable. *Behaviour Research and Therapy, 27*(3), 279-287.

Baeyens, F., Eelen, P., Van den Bergh, O., & Crombez, G. (1990). Flavor-flavor and color-flavor conditioning in humans. *Learning and Motivation, 21*(4), 434-455.

Baeyens, F., Hendrickx, H., Crombez, G., & Hermans, D. (1998). Neither extended sequential nor simultaneous feature positive training result in modulation of evaluative flavor-flavor conditioning in humans. *Appetite, 31*(2), 185-204.

Bar-Anan, Y., De Houwer, J., & Nosek, B. A. (2009). *The role of contingency memory, intentional processes and number of pairings in evaluative conditioning*. Unpublished Manuscript.

Bouton, M. E. (2002). Context, ambiguity, and unlearning: Sources of relapse after behavioral extinction. *Biological Psychiatry, 52*(10), 976-986.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11*(5), 485-494.

Cacioppo, J. T., Marshall-Goodell, B. S., Tassinary, L. G., & Petty, R. E. (1992). Rudimentary determinants of attitudes - Classical-conditioning is more effective when prior knowledge about the attitude stimulus is low than high. *Journal of Experimental Social Psychology, 28*(3), 207-233.

Coppens, E., Vansteenwegen, D., Baeyens, F., Vandenbulcke, M., Van Paesschen, W., & Eelen, P. (2006). Evaluative conditioning is intact after unilateral resection of the anterior temporal lobe in humans. *Neuropsychologia, 44*(5), 840-843.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measurement: Consistency, stability, and convergent validity. *Psychological Science, 12*, 163-170.

Cuthbert, B. N., Bradley, M. M., & Lang, P. J. (1996). Probing picture perception: Activation and emotion. *Psychophysiology, 33*(2), 103-111.

Dawson, M. E., Beers, J. R., Schell, A. M., & Kelly, A. (1982). Allocation of cognitive processing capacity during human autonomic classical-conditioning. *Journal of Experimental Psychology - General, 111*(3), 273-295.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology, 10*, 230-241.

De Houwer, J. (2009). Conditioning as a source of liking: There is nothing simple about it. In M. Wänke (Ed.). *Frontiers of Social Psychology: The Social Psychology of Consumer Behavior.* New York: Psychology Press.

De Houwer, J. (in press). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds.), *Applications of learning and conditioning*. Oxford, UK: Oxford University Press.

De Houwer, J., Baeyens, F., & Field, A. P. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*(2), 161-174.

De Jong, P. J., Vorage, I., & van den Hout, M. A. (2000). Counterconditioning in the treatment of spider phobia: Effects on disgust, fear and valence. *Behaviour Research and Therapy, 38*(11), 1055-1069.

Diaz, E., Ruiz, G., & Baeyens, F. (2005). Resistance to extinction of human evaluative conditioning using a between-subjects design. *Cognition & Emotion, 19*, 245-268.

Dirikx, T., Hermans, D., Vansteenwegen, D., Baeyens, F., & Eelen, P. (2004). Reinstatement of extinguished conditioned responses and negative stimulus valence as a pathway to return of fear in humans. *Learning & Memory, 11*(5), 549-554.

Field, A. P. (2000). Evaluative conditioning is pavlovian conditioning: Issues of definition, measurement, and the theoretical importance of contingency awareness. *Consciousness and Cognition, 9*(1), 41-49.

Field, A. P., & Davey, G. C. L. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology - Animal Behavior Processes, 25*(2), 211-224.

Field, A. P., Lascelles, K. R. R., Lester, K. J., Askew, C., & Davey, G. C. L. (2008). Evaluative conditioning: Missing, presumed dead. *Netherlands Journal of Psychology, 64*, 46-64.

Fulcher, E. P., & Hammerl, M. (2001). When all is revealed: A dissociation between evaluative learning and contingency awareness. *Consciousness and Cognition, 10*(4), 524-549.

Hamm, A. O., & Vaitl, D. (1996). Affective learning: Awareness and aversion. *Psychophysiology, 33*(6), 698-710.

Hardwick, S. A., & Lipp, O. V. (2000). Modulation of affective learning: An occasion for evaluative conditioning? *Learning and Motivation, 31*(3), 251-271.

Hermans, D., Baeyens, F., Lamote, S., Spruyt, A., & Eelen, P. (2005). Affective priming as an indirect measure of food preferences acquired through odor conditioning. *Experimental Psychology, 52*(3), 180-186.

Hermans, D., Spruyt, A., & Eelen, P. (2003). Automatic affective priming of recently acquired stimulus valence: Priming at SOA 300 but not at SOA 1000. *Cognition & Emotion, 17*(1), 83-99.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (in press). Evaluative conditioning in humans: A meta-analysis. *Psychological Bulletin*.

Kerkhof, I. (2006). *Counterconditioning in a differential fear paradigm*. Unpublished data.

Kerkhof, I. (2008). *Counterconditioning in a picture-picture paradigm*. Unpublished data.

Korrelboom, C. W., van der Gaag, M., Hendriks, V. M., Huijbrechts, I. P. A. M., & Berretty, E. W. (2008). Treating obsessions with Competitive Memory Training: A pilot study. *The Behavior Therapist, 31*, 29-36.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2005). *International affective picture system (IAPS): Digitized photographs, instruction manual and affective ratings. Technical Report A-6.* University of Florida, Gainesville, FL.

Lascelles, K. R. R., & Davey, G. C. L. (2006). Successful differential evaluative conditioning using simultaneous and trace conditioning procedures in the picture-picture paradigm. *The Quarterly Journal of Experimental Psychology, 29*, 482-492.

Lascelles, K. R. R., Field, A. P., & Davey, G. C. L. (2003). Using foods as CSs and body shapes as UCSs: A putative role for associative learning in the development of eating disorders. *Behavior Therapy, 34*(2), 213-235.

Lipp, O. V., Oughton, N., & Lelievre, J. (2003). Evaluative learning in human pavlovian conditioning: Extinct, but still there? *Learning and Motivation, 34*(3), 219-239.

Lipp, O. V., & Purkis, H. M. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & Emotion, 19,* 269-282.

Lipp, O. V., & Purkis, H. M. (2006). The effects of assessment type on verbal ratings of conditional stimulus valence and contingency judgments: Implications for the extinction of evaluative learning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(4), 431-440.

Lipp, O. V., Siddle, D. A. T., & Dall, P. J. (1993). Effects of miscuing on pavlovian conditioned responding and on probe reaction-time. *Australian Journal of Psychology, 45*(3), 161-167.

Martin, I., & Levey, A. (1994). The evaluative response - Primitive but necessary. *Behaviour Research and Therapy, 32*(3), 301-305.

Nairne, J. S., & Pandeirada, J. N. S. (2008). Adaptive memory: Is survival processing special? *Journal of Memory and Language, 59*, 377-385.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology - General, 106*, 226-254.

Payne, B. K., Cheng, C. M., Govorun, O., & Stewart, B. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89*, 277-293.

Peck, C. A., & Bouton, M. E. (1990). Context and performance in aversive-to-appetitive and appetitive-to-aversive transfer. *Learning and Motivation, 21*(1), 1-31.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis)liking: Item-based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology - Learning Memory and Cognition, 33*(1), 130-144.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance: V* (pp. 669-682). New York: Academic Press.

Rozin, P., Wrzesniewski, A., & Byrnes, D. (1998). The elusiveness of evaluative conditioning. *Learning and Motivation, 29*(4), 397-415.

Shanks, D. R., & St. John, M. F. (1994). Characteristics of dissociable human learning-systems. *Behavioral and Brain Sciences, 17*(3), 367-395.

Spruyt, A., De Houwer, J., & Hermans, D. (2009). Modulation of automatic semantic priming by feature-specific attention allocation *Journal of Memory and Language, 61,* 37-54.

Stevenson, R. J., Boakes, R. A., & Wilson, J. P. (2000). Counter-conditioning following human odor-taste and color-taste learning. *Learning and Motivation, 31*(2), 114-127.

Stuart, E. W., Shimp, T. A., & Engle, R. W. (1987). Classical-conditioning of consumer attitudes - 4 experiments in an advertising context. *Journal of Consumer Research, 14*(3), 334-349.

Todrank, J., Byrnes, D., Wrzesniewski, A., & Rozin, P. (1995). Odors can change preferences for people in photographs - A cross-modal evaluative conditioning study with olfactory USs and visual CSs. *Learning and Motivation, 26*(2), 116-140.

Vansteenwegen, D., Francken, G., Vervliet, B., Declercq, A., & Eelen, P. (2006). Resistance to extinction in evaluative conditioning. *Journal of Experimental Psychology - Animal Behavior Processes, 32*, 71-79.

Vansteenwegen, D., Hermans, D., Vervliet, B., Francken, G., Beckers, T., Baeyens, F., & Eelen, P. (2005). Return of fear in a human differential conditioning paradigm caused by a return to the original acquistion context. *Behaviour Research and Therapy, 43*(3), 323-336.

Verhulst, F., Hermans, D., Baeyens, F., Spruyt, A., & Eelen, P. (2006). Determinants and predictive validity of direct and indirect measures of recently acquired food attitudes. *Appetite, 46*(2), 137-143.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009). Changing likes and dislikes through the back door: The US-revaluation effect. *Cognition & Emotion, 23*(5), 889-917.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four-picture recognition test. *Journal of Experimental Psychology - Animal Behavior Processes, 32*(4), 454-459.

Wardle, S. G., Mitchell, C. J., & Lovibond, P. E. (2007). Flavor evaluative conditioning and contingency awareness. *Learning & Behavior, 35*(4), 233-241.